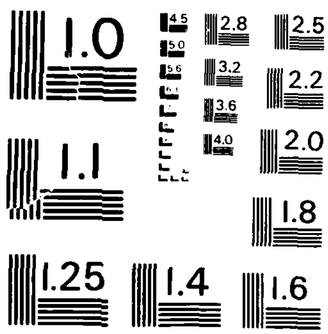
AD-A163 833 A COMPARISON OF ESTIMATION TECHNIQUES FOR THE TWO PARAMETER CAUCHY DISTRIBUTION(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGI.. J O SOURS 1/2

UNCLASSIFIED DEC 85 AFIT/GSO/MA/85D-7 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A163 833

DTIC
ELECTE
FEB 1 0 1986
S D
D

A COMPARISON OF ESTIMATION

TECHNIQUES FOR THE TWO

PARAMETER CAUCHY DISTRIBUTION

THESIS

John O. Sours
Captain, USAF

AFIT/GSO/MA/85D-7

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GSO/MA/85D-7

(1)

A COMPARISON OF ESTIMATION

TECHNIQUES FOR THE TWO

PARAMETER CAUCHY DISTRIBUTION

THESIS

John O. Sours
Captain, USAF

AFIT/GSO/MA/85D-7

AFIT/GSO/MA/85D-7

A COMPARISON OF ESTIMATION TECHNIQUES FOR THE TWO PARAMETER CAUCHY DISTRIBUTION

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Accesion For	
NTIS CRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

by

John O. Sours, B.A.
Captain USAF

Graduate Space Operations

December 1985

Approved for public release; distribution unlimited

Acknowlegments

I am extremely indebted to Dr. Albert Moore for suggesting this research topic and directing my research effort. I truly enjoyed working with him and appreciated his effort in stimulating my interest in this area. I would also like to thank my reader, Dr. Joseph Cain, whom I chose because of his previous experience in past minimum distance efforts. I found his comments to be most helpful. Thanks must also go to Major Denny Charek, Capt Jim Porter, and Capt Frank Ocasio who did similar research in this area and with whom I exchanged much information that proved to be very useful.

Finally, I would like to thank my family, who were patient enough to endure my stay here at AFIT. This includes my precious wife Juanita, without whose encouragement, sacrifice, and help I could have never accomplished this work, my son Brad, who had to endure not having a father to play with at times, and my daughter Hilary, who was born during this tour. Last but not least, I would like to thank the Lord Jesus Christ, my personal Lord and Savior, for guiding me through this thesis effort. It is written in Psalms 32:8: "I will instruct you and guide you along the best pathway for your life". Truly without his guidance, I doubt that I could have made the effort at all.

John O. Sours

Table of Contents

		Page
Acknowledgments		ii
List of Figures		v
List of Tables		vi
Abstract		vii
I.	Introduction	1
	Background	1
	Problem Statement	4
	Research Question	5
	Methodology	5
	Chronology	6
II.	The Cauchy Distribution	7
	Introduction.	7
	Probability Density Function	7
	Cumulative Distribution Function	10
	Stable Law and the Characteristic Function.	10
	Important Statistics	11
	Order Statistics and Censored Samples	12
	Mathematical Applications	12
III.	Parameter Estimation of the Cauchy Distribution	14
	Introduction.	14
	Estimators of the Cauchy Distribution	14
	Maximum Likelihood Estimator	16
IV.	Minimum Distance Estimation	19
	Introduction.	19
	Defintions and Notation	20
	Weighted Kolmogorov Distance	22
	Weighted Cramer-von Mises Distance	23
	Anderson-Darling Statistic	24
	Kuiper Distance.	25

Watson Distance 26
Empirical Distribution Function Distance . 28
Summary 28

V. Monte Carlo Analysis 30

Introduction. 30
Generation of Data. 30
Initial Estimates 31
Maximum Likelihood Estimates 31
Minimum Distance Estimators. 32
Kolmogorov Estimator 34
Cramer von-Mises Estimator 34
Anderson-Darling Estimator 35
Kuiper Estimator 35
Watson Estimator 36
Evaluation Criteria 36
A Walk-Through the Computer Program . . . 37

VI. Conclusions and Recommendations 40

Results 40
Conclusions 42
Follow-on Work 43

Bibliography 45

Appendix A: Tables. 48

Appendix B: Computer Program Listing of CLPEE. . . 62

Appendix C: Computer Program Listing of CSPEE. . . 76

Appendix D: Computer Subprogram Li ng of VEDIS and
AEDIS for Censored Samples 87

VITA. 90

List of Figures

Figure | Page

1. Normal versus Cauchy 8
2. Geometric Represention of Cauchy Distributed Impact Points 13
3. CDF versus EDF 21
4. Kuiper Distance. 27
5. Flow Chart of Cauchy Location Parameter Estimation and Efficiencies 39

List of Tables

Table		Page
I.	Non-Converging MLE Estimates (1000 Samples)	48
II.	MSE and RE of Estimators (Tables 1-12)	
1.	Non-Censored Samples, Sample Size = 6	49
2.	Non-Censored Samples, Sample Size = 8	50
3.	Non-Censored Samples, Sample Size = 10	51
4.	Non-Censored Samples, Sample Size = 12	52
5.	Non-Censored Samples, Sample Size = 16	53
6.	Censored Samples, Sample Size = 6	54
7.	Censored Samples, Sample Size = 8	55
8.	Censored Samples, Sample Size = 10	56
9.	Censored Samples, Sample Size = 12	57
10.	Censored Samples, Sample Size = 16	58
11.	Location = -2, Scale = 5, Sample Size = 6	59
12.	Location = -2, Scale = 5, Sample Size = 16	60
III.	MSE and RE of Watson Estimators	61

Abstract

Minimum distance estimators are compared to the maximum likelihood estimator of the location and scale parameters of the two parameter Cauchy distribution. Samples sizes of 6,8,10,12, and 16 are randomly drawn from a Cauchy distribution and used to estimate the parameters by the maximum likelihood method. Minimum distance techniques are then used to try improve upon the maximum likelihood estimates. 1000 samples for each sample size are generated and the mean squared error and relative efficiencies are computed. Comparison of the two methods is done by comparing the relative efficiencies. Both censored and non-censored samples are used to compute the minimum distance estimates.

COMPARISON OF ESTIMATION TECHNIQUES FOR THE TWO PARAMETER CAUCHY DISTRIBUTION

I. Introduction

Background

Military decision makers have increasingly been using information obtained from statistical analysis in the decision making process. This information, which can include description of data, interpretation of data, and point and confidence interval estimation (7:93), allows decision makers to be better informed and to make better decisions. Two areas in which military decision makers have extensively used statistical analysis include the reliability of weapons systems and simulation of military systems and scenarios.

For example, consider the computer simulation of a nuclear war between the Soviet Union and the United States. This situation could be modeled on a computer using a simulation language such as SLAM. Statistical analysis could be used to estimate the mean distance of a missile warhead from a target point. This estimate could be derived from a series of test flights where the distances between impact points and target points are recorded. This estimation could then be used as a parameter in a probability distribution which generates individual warheads at various impact points over a period of time. Other distributions such as the probability of detonation of a particular warhead could

be derived in a similar manner using statistical analysis. These and other factors could then be used to simulate a real world nuclear engagement. From this simulation, inferences could be drawn about a real world nuclear engagement and decision makers would be better informed to make decisions concerning weapon types, numbers, deployment, etc. Clearly, the more accurate the parameter estimation, the more reliable the simulation, and hense the better the information available to the decision maker. Parameter estimation is therefore a critical step in the decision making process.

The beginning of parameter estimation began in the early 1800s when Legrende and Gauss independently derived the method of least squares. Lengrende used the method to estimate the orbital parameters of comets and as this author well knows, the method is still used today to estimate the orbital parameters of artificial satellites. Gauss gave the probabilistic basis for the method and worked out the computational techniques (22:14). The first major advance in the theory occurred in the late 1800s when the method of moments was formulated by K. Pearson. This method is considered one of the classical estimation techniques although it is not as efficient as other methods used today. It is usually employed when other estimators are difficult to compute (10:6). The present day foundation of estimation theory was developed by R.A. Fisher in the 1920s. His method, called the method of maximum likelihood is usually superior to the

method of moments. Gauss had actually anticipated the maximum likelihood method, but felt that it was inferior to the method of least squares (10:6).

A more recent method of parameter estimation is the minimum distance method. This method was developed by Wolfowitz in the 1950s under a contract with the Department of Defense. The method assumes a probability distribution, calculates an initial estimate of the parameter by another method, and then tries to improve that estimate by minimizing the distance between the empirical distribution and the assumed or theoretical distribtion. Wolfowitz states that "the minimum distance method will, in a wide variety of cases, furnish super-consistent estimators even when the classical methods, such as the maximim likelihood methods, fail to give consistent estimators" (29:9,30:203). Distance estimation techniques for various distributions have also generally shown improvement over the maximum likelihood estimators (4:9).

Comparison of the minimum distance estimator with other estimators has been an effort guided by Dr Albert H. Moore, a professor of statistics at the Air Force Institute of Technology. He has directed several thesis efforts in estimating distribution parameters by the minimum distance technique and comparing these estimations with other estimators. These thesis efforts have included comparing estimators of the four-parameter beta distribution, the exponential distribution, the four-parameter generalized gamma dis-

tribution, the generalized t distribution, the three parameter Weibull distribution, and the three parameter lognormal distribution. An unpublished thesis concurrently compared minimum distance estimators with other estimators of the three parameter Pareto distribution.

Another distribution in which minimum distance estimation has not yet been fully explored is the Cauchy distribution. The Cauchy, named after the French mathematician Augustus Cauchy, is a probability distribution which has been largely ignored in modeling because the expected value of its parameters are not well defined and are difficult to estimate accurately by conventional methods. The distribution has received more attention lately, however, due to the computer's ability to quickly compute estimates of its parameters. The Cauchy has been shown useful in mechanics, electrical theory, and measurement and calibration problems (9:418), and has lately been considered as an alternative to the normal distribution in describing error distributions where the error terms have very long tails (13:114,23:9). This might find application in simulation studies (26:2). Also, infinite variance distributions (such as the Cauchy) have been considered in time series analyses of economic data (11:275).

Problem Statement

Minimum distance estimation has not been applied to estimators of the location or scale parameter of the Cauchy

distribution. A controlled experiment needs to be performed to determine if minimum distance can improve upon estimators such as the maximum likelihood estimator (mle).

Research Question

Will minimum distance estimation improve the estimation of the location/scale parameter over that of the mle? If the estimation of the location/scale parameter is improved, will the new estimate improve the maximum likelihood estimate of the scale/location parameter? How much will the minimum distance technique improve the estimates? Will using censored samples (samples with certain data points removed) make a difference in the estimates?

Methodology

The research questions will be answered by setting up a controlled environment and performing a Monte Carlo analysis. First, Cauchy random variates will be generated from a Cauchy distribution. The Cauchy random variate generator GGCAY found in the International Mathematical Statistics Library (IMSL) will be utilized to generate the necessary samples. Sample sizes of 6,8,10,12, and 16 will be generated with a location parameter of 0 and a scale parameter of 1. Both censored and non-censored samples will be used. A second set of samples with sample sizes of 6 and 16 will be generated with a location parameter of -2 and a scale parameter of 5. This variation in the location and scale parameters will allow investigation of the invariance of the

estimators.

Once the necessary samples are generated, the location and scale parameters will be estimated by the mle method. Depending on the distance measure used, the minimum distance technique will try to improve the mle of the location/scale parameter. Both estimates will be compared to the original parameter value to see which one is better. If the location/scale estimate is improved, a new scale/location parameter will be estimated using the mle. This new estimate will be compared to the original mle of the scale/location parameter to see if there is any improvement. Finally, the degree of improvement will be calculated for both parameters and the results tabulated.

Chronology

Chapter Two will describe the Cauchy distribution. Chapter Three will discuss estimators of the Cauchy distribution with emphasis on the maximum likelihood estimator. Chapter Four will discuss minimum distance estimation and describe the different distance estimators including estimators using censored samples. Chapter Five will discuss the Monte Carlo analysis, describe how to compute the distance measures, discuss the programming involved, and define the criteria by which to evaluate the methods. Chapter Six will discuss results, conclusions, and suggest follow-on work.

II. THE CAUCHY DISTRIBUTION

Introduction

The Cauchy is a special form of a general class of distributions called the Pearson Type VII distributions (16:154) which are unimodal and symetric about a central location parameter. They have the functional form

$$dF = (1/k\psi)(\ 1 + ((x-\lambda)/\psi)^2)^{-m} dx \qquad 2\text{-}1$$

where $m > 1/2$

and k is a constant (19:302).

The location parameter, λ , serves both as the unique median and mode. The distribution reduces to the Cauchy when m equals 1 and k equals π, and looks very similar to the normal as shown in Figure 1, although it is more peaked and thicker tailed. Because of the Cauchy's infinite variance, there is no standardized form of the distribution, however, the Cauchy does have a standard form by setting the location parameter to 0 and the scale parameter, ψ , to 1. When this is done, the form is the same as the t-distribution with one degree of freedom.

Probability Density Function

The probability density function (pdf) of the Cauchy has the form

$$f(x) = 1/(\pi(\ 1 + ((x-\lambda)/\psi)^2) \qquad 2\text{-}2$$

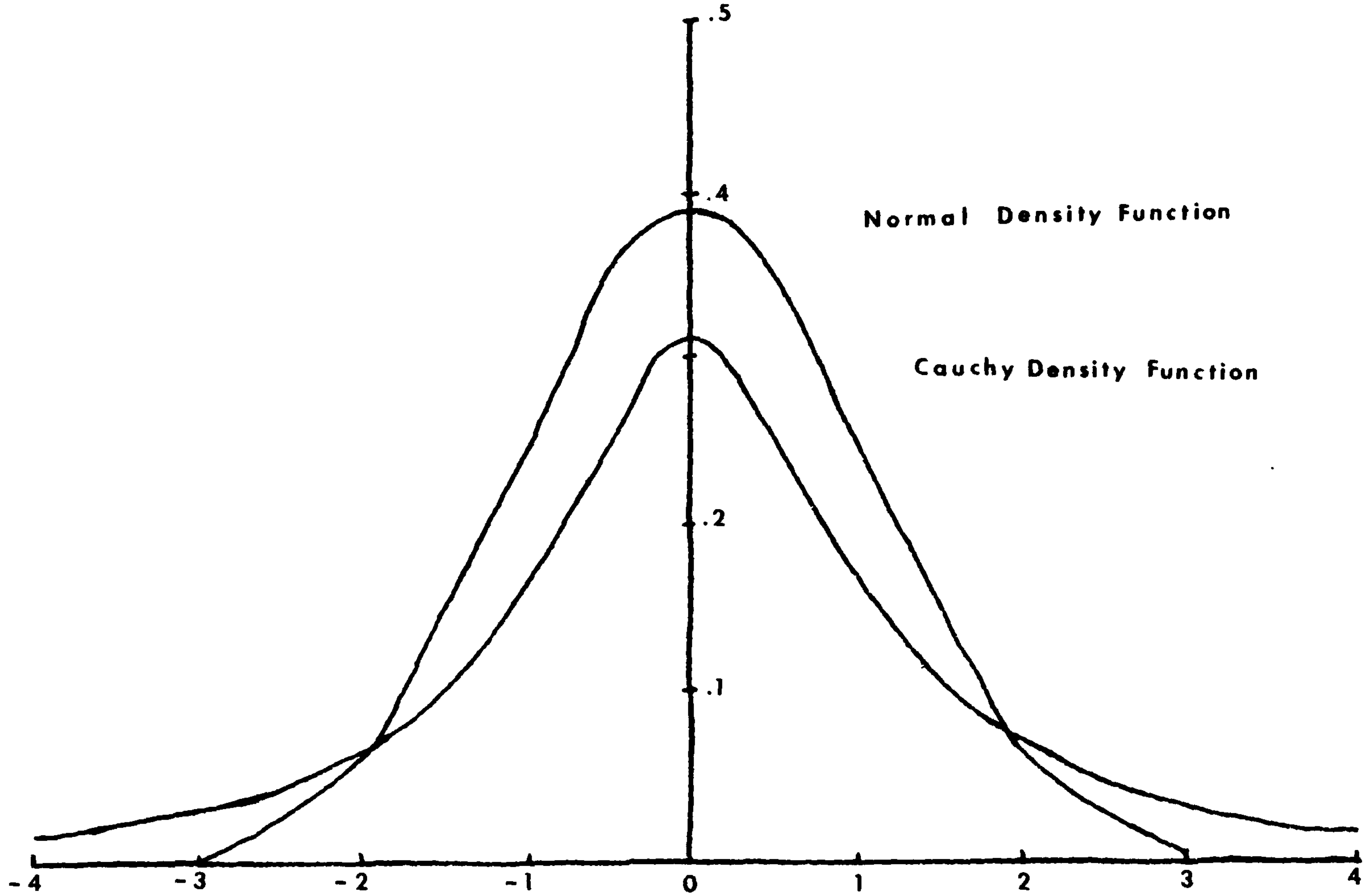


Figure 1. Normal versus Cauchy

with $(-\infty < x < +\infty)$

$(-\infty < \lambda < +\infty)$

$(\psi > 0)$

where λ , as already mentioned, is the location parameter and ψ is the scale parameter. The distribution is symetric about the median λ with the upper and lower quartiles equal to $\lambda \pm \psi$.

One reason why the Cauchy has not been used in common statistical applications is because the statistical theory that exists for the normal is nonexistent for the Cauchy (11:275). This is partially because the expected value (E(x)) of the location parameter (also known as the first moment) is not well defined. Since E(x) is defined as the integral of xf(x) and as b approaches $+\infty$ and a approaches $-\infty$,

$$\int_a^b x(1/(\pi(1+((x-\lambda)/\psi)^2))dx =$$

$$(1/2\pi)[\log(1+b^2)-\log(1+a^2)] \qquad 2\text{-}3$$

which is infinite, the expected value of location is not well defined (32:205) (Note: In the above integral, $\psi = 1$ and $\lambda = 0$. The result would be the same if the parameters held different values). Therefore the mean value of a Cauchy sample cannot represent the expected value of the location parameter and regardless of how large the sample is, the variability of the mean about a central value does not decrease as the sample size increases (9:417). Therefore, the sample mean is not a useful estimate of the loca-

tion parameter.

The other problem in using the Cauchy in statistical applications lies in the Cauchy's infinite variance (also known as the second moment). This is also a direct result of (2-3). This can create profound problems because as Granger reports:

> alternative measures of dispersion such as absolute deviations are much harder to work with, and in a regime of infinite variance, the squares of deviations become suspect as a measure of dispersion, since the means of squares diverge in the non-Gaussian populations (11:276).

Other estimates of the parameters must therefore be used. With the aid of the computer, however, many estimates can be easily evaluated.

Cumulative Distribution Function

Integrating (2) from $-\infty$ to $+\infty$ yields the cumulative distribution function (cdf) which has the form

$$F(x) = 1/2 + (1/\pi)\arctan\{(x-\lambda)/\psi\} \qquad 2\text{-}4$$

Stable Law and the Characteristic Function

The Cauchy distribution follows stable law which implies that the sum of two Cauchy variates is also Cauchy distributed (19:273). Stable distributions are usually defined by their characteristic function (20:918) because very few can be written explicitly (11:275). The Cauchy, however, can be written explicitly as (2-2).

The Cauchy empirical characteristic function has the form

$$Cn(t) = \int_{-\infty}^{\infty} e^{itz}\, dGn(z) \;, \quad -\infty < t < +\infty \qquad 2\text{-}5$$

as described by Parr (25:1207).

Important Statistics

Because the mean and standard deviation of the Cauchy sample are not useful in estimating the location and scale parameters, other estimates must be used. Two easily computed estimators are the median for the location parameter and the semi-interquartile range for the scale.

The median is defined as follows. After obtaining a Cauchy sample of size N, the independently random variates Xi of the sample are ordered such that $X1 \leq X2 \leq \ldots\ldots Xn$. The median is then defined as the .5 quantile of the sample or

$$\text{median} = X_{.5}\,.$$

If the sample size is odd, the median will be the value of the middle variate or $X_{(\frac{n+1}{2})}$. If the sample size is even, the median will be the average of the middle two variates or

$$X = (1/2)\left(X_{\frac{n}{2}} + X_{(\frac{n+2}{2})} \right).$$

The semi-interquartile range is simply defined as the sum of the average of the .25 and .75 quantiles (9:216) or

$$\text{semi-interquartile range} = (1/2)\left(X_{.25} + X_{.75} \right).$$

In this thesis, a modified semi-interquartile range is used. The initial starting point is determined by first computing the 25% and 75% quantiles. The distance from the median to each quantile is then computed and the smaller of these distances is doubled and used as the estimate.

Order Statistics and Censored Samples

If a random sample of independent random variables $X1, X2, \ldots. Xn$ are ordered such that $X'1 \leq X'2 \leq \ldots.. X'n$, then the ordered variables are called the corresponding order statistics (16:157) and each order statistic from the Cauchy distribution has a probability density function of

$$P_{X'_r}(x) = (n!/((r-1)!(n-r)!))(1/2 + (1/\pi)\arctan\{(x-\lambda)/\psi\}) \times$$

$$(1/2 - (1/\pi)\arctan\{(x-\lambda)/\psi\}) \quad (1/(\pi\psi))(1 + \{(x-\lambda)/\psi\}) \qquad 2\text{-}6$$

The expected values the first and last order statistics and the variances of the second and next-to-last order statistics are infinite for Cauchy ordered samples (16:157). The sample is said to be censored when the first two and last two variates of the ordered sample are not used in estimating the Cauchy parameters. These variates are called the censored variates.

Mathematical Applications

If U and V are independent unit normal variables, then the distribution of U/V has a standard Cauchy distribution (16:160). The common distribution of U and V does not have

to be normal. If c and d are nonzero constants and X is Cauchy distributed, then cX + dX is also Cauchy distributed. The reciprocal of a Cauchy variable has a Cauchy distribution as well.

Geometrically, the Cauchy represents the diagram in Figure 2. If the angle OAP is uniformly distributed between $-\pi/2$ to $\pi/2$, the distance OP is Cauchy distributed. This situation might arise if a particle is emitted from a source point A and travels in a straight line to an impact point P on a line perpendicular to AO and in the same plane as angle OAP. The angle $-\pi/2$ represents $-\infty$ and the angle $\pi/2$ represents $+\infty$. The concept can be expanded to three dimensions as well.

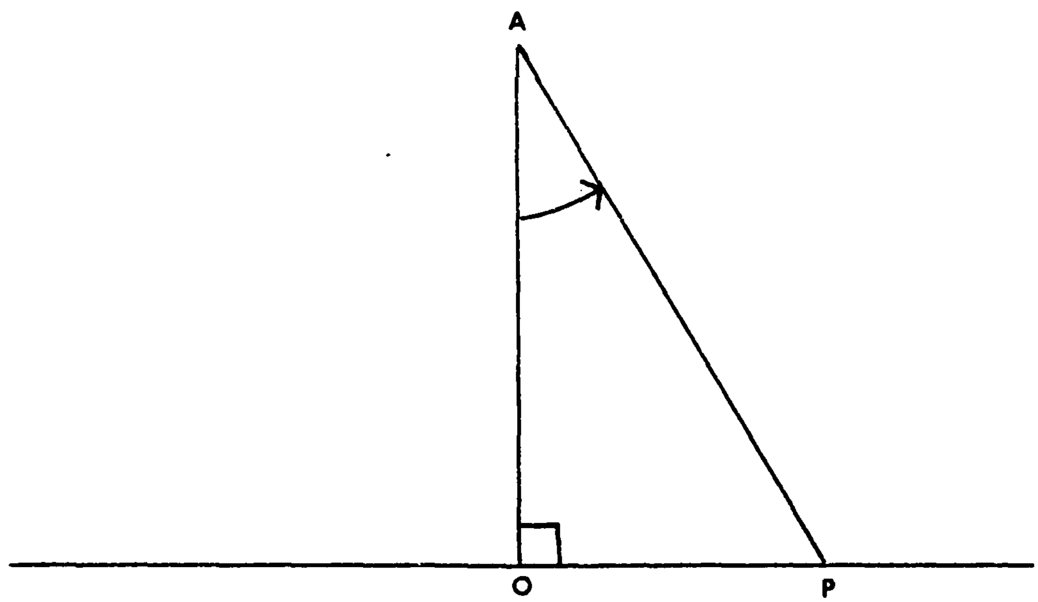


Figure 2. Geometric Representation of Cauchy Distributed Impact Points

III. Parameter Estimation of the Cauchy Distribution

Introduction

Parameter estimation was introduced in Chapter One. This chapter will continue the discussion of parameter estimation with specific reference to the Cauchy distibution. It should be noted that the term "efficiency" in the following discussion refers to the accuracy of an estimator or how close the estimator actually approaches the real value of the parameter.

Estimators of the Cauchy Distribution

There are several ways to estimate the parameters of the Cauchy distribution. One of the more simpler and reliable estimates of location is the median (1:240) which is often used along with the semi-interquartile range as initial estimates for iterative methods. Two other more efficient methods (2:1205-1206) include the use of order statistics and the maximum likelihood estimator (mle). The method of moments, which can be used with many other distributions, does not apply to the Cauchy because as shown in Chapter Two, the moments of the Cauchy do not exist.

The literature search revealed several methods which rely on the use of order statistics. One the most efficient order statistic methods is the Best Linear Unbiased Estimator (BLUE) proposed by Barnett in 1966. The BLUE involves extensive calculations of the variances and covariances of the order statistics. These values have been calculated and

tabulated for samples of size n=5(1)16(2)20 and achieve full asymptotic efficiency and small sample efficiencies of at least 80% when compared to the mle (2:1206). Barnett says that the evaluation of the estimate with the aid of tabulated coefficients is relatively simple when compared to the mle. Other order statistics methods, although easier to compute, are less efficient than the BLUE (21:211-212).

Since the mle does not exist in closed form, it must be computed numerically (13:114). Barnett reported a potential problem of multiple solutions in solving the likelihood equations but Haas, et al, reported that while actually sampling from the Cauchy distribution, multiple solutions were never found (12:404). They conjectured that the solution of the likelihood equations will always be unique for distinct sample sizes of three or more and Copas later showed this to be true (8:701). Thus the difficulties in computing the mle as forseen by Barnett are not as formidable as previously thought.

A more recent method of estimating the Cauchy parameters is through the use of window estimates proposed by Higgins and Tichenor (14:157). This is a closed form method and has the same asymptotic distributions as the maximum likelihood estimates (13:114). These estimates have high efficiencies along with the maximum likelihood estimates for moderate to large sample sizes (14:164). Smaller sample sizes, however, are less efficient. For example, the variance of the estimate of the location parameter by window

estimation is larger than that of a similar estimate by the mle for a sample size of 10 (14:164). Higgins does report, however, that a major advantage of his method is its ease of computation (13:113).

Another recent estimation technique involves order statistics and use of the empirical characteristic function (21:205). The efficiencies of these estimates were found to be superior to the BLUE but the author did not compare his estimates to that of the mle. Observing the variance of the characteristic function estimator for a sample size of 10 reveals that this variance is larger than that of the mle.

The mle, in all cases considered, proved to be the most efficient estimator. For this reason, the mle was chosen to be the initial estimate of the Cauchy parameters before applying minimum distance techniques.

Maximum Likelihood Estimator

The maximum likelihood function, $L(X1,X2,...Xn;\lambda,\psi)$, for a sample size n from the Cauchy distribution is given by Haas, et al, as

$$L(X1,X2,...Xn;\lambda,\psi) = \prod_{i=1}^{n} [\ 1/(\pi\psi(1 + (Xi-\lambda\ /\psi)^2))] \qquad 3\text{-}1$$

with the logarithm of the likelihood function as

$$\text{Log } L = -n(\log(\pi)) - n(\log(\lambda)) - \sum_{i=1}^{n} \log\{1+((Xi-\lambda)/\psi)^2\}. \qquad 3\text{-}2$$

Hense the maximum likelihood equation are

$$\sum_{i=1}^{n} ((Xi-\lambda)/\psi)/\{1 + ((Xi-\lambda)/\psi)^2\} = 0 \qquad 3\text{-}3$$

$$\sum_{i=1}^{n} 1/\{1 + ((Xi-\lambda)/\psi)^2)\} = 1/(2n) \qquad 3\text{-}4$$

where λ and ψ are the estimates of the location and scale respectively. The equations must must be solved numerically (12:404).

Haas used the Newton-Rhapson method to solve the equations but they can also be solved by an iterative method. This was the method used by the Princeton Study (1:16) and the iterative solutions are shown in equations 3-5 and 3-6. Equation 3-5 is used to estimate the location parameter while equation 3-6 is used to estimate the scale.

$$\lambda_{k+1} = \sum_{i=1}^{n} (Xi/[\psi_k^2 + (Xi-\lambda_k)^2])/\sum_{i=1}^{n} (1/[\psi_k + (Xi-\lambda_k)\ ]) \qquad 3\text{-}5$$

$$\sqrt{\psi_{k+1}} = n/[\,2(\psi_k)^{\frac{3}{2}} \sum_{i=1}^{n} 1/(\psi_k^2 + (Xi-\lambda_k)^2)]. \qquad 3\text{-}6$$

Both equations require initial starting values to initiate the iteration. The median is used as the starting point for the location and is designated by λ_k. The modified semi-interquartile range is used as the starting point for the scale and is designated by ψ_k. This was the estimate used in the Princeton study. Each succeeding iteration uses the most previous iterated values of location and scale as initial starting values.

Both equations 3-5 and 3-6 were incorporated into a Fortran subroutine called CMLE (Cauchy Maximum Likelihood Estimator) to generate the estimates. The routine was adapted from the Fortran function listed in the Princeton Study (1:262). A listing of the subroutine can be found in the main program CLPEE in Appendix B. Results of the estimates are discussed in Chapter Five.

IV. Minimum Distance Estimation

Introduction

The minimum distance (MD) method of parameter estimation was pioneered by Wolfowitz during the 1950s. In his fundamental paper (1957), he outlined the MD method, gave a number of examples concerning its use, and proved consistency of the estimators (converging with a probability of one) (24:616,31:75). A decade later, other work followed including Sahler (1970), who surveyed a variety of empirical distributions, outlined conditions for the existence and consistency of MD estimators, and proved an asymptotic normality result (18:33). Knusel (1969) examined the robustness of the MD method and showed that MD estimators have similar robust properties as the maximum likelihood estimators. Robustness, here, refers to the ability of an estimator to be reasonably accurate even though the underlying distribution deviates somewhat from the theoretical distribution. Parr and Schucany also investigated the robustness properties of MD estimators and used MD techniques to estimate the location parameter of the normal distribution. They used samples from a wide variety of symetric distributions including the Cauchy. Other leaders in the field of MD estimation include Blackman, Rao, and Littlell. For a complete list of contributions on MD estimators, see Parr's bibliography on minimum distance estimation (23).

Definition and Notation

Basically the MD method involves minimizing the distance or discrepency between two cumulative distribution functions. One function is the cumulative empirical distribution function (edf) which is determined from an ordered random sample $X1, X2, ... Xn$ where $X1 \leq X2 \leq .. Xn$. The cumulative edf is defined as the function which equals the fraction of Xis that are less than or equal to x for each x, $-\infty < x < +\infty$ (7:69). The function is a step function and can be written as

$$Gn(x) = i/n \tag{4-1}$$

where i is simply the position number of the Xi point. For all x in the interval $Xi \leq x < X_{i+1}$, $Gn(x)$ is equal to i/n. The second function is the assumed cumulative probability distribution and comes from a parameterized family of theoretical distribution functions. $F_\theta(x;\theta\varepsilon\Omega)$ represents the theoretical funcion where θ is a particular parameter and Ω represents the parameter space. In this thesis, F_θ represents the Cauchy distribution with estimated scale and location parameters and Gn represents the edf based on the random sample. Figure 3 compares the cdf for both the edf and theoretical distribution.

The discrepency between the two is measured. The parameter being estimated is then altered by a small value and a new discrepency is calculated. The new discrepency is compared with the old one and the one which has the mimimum

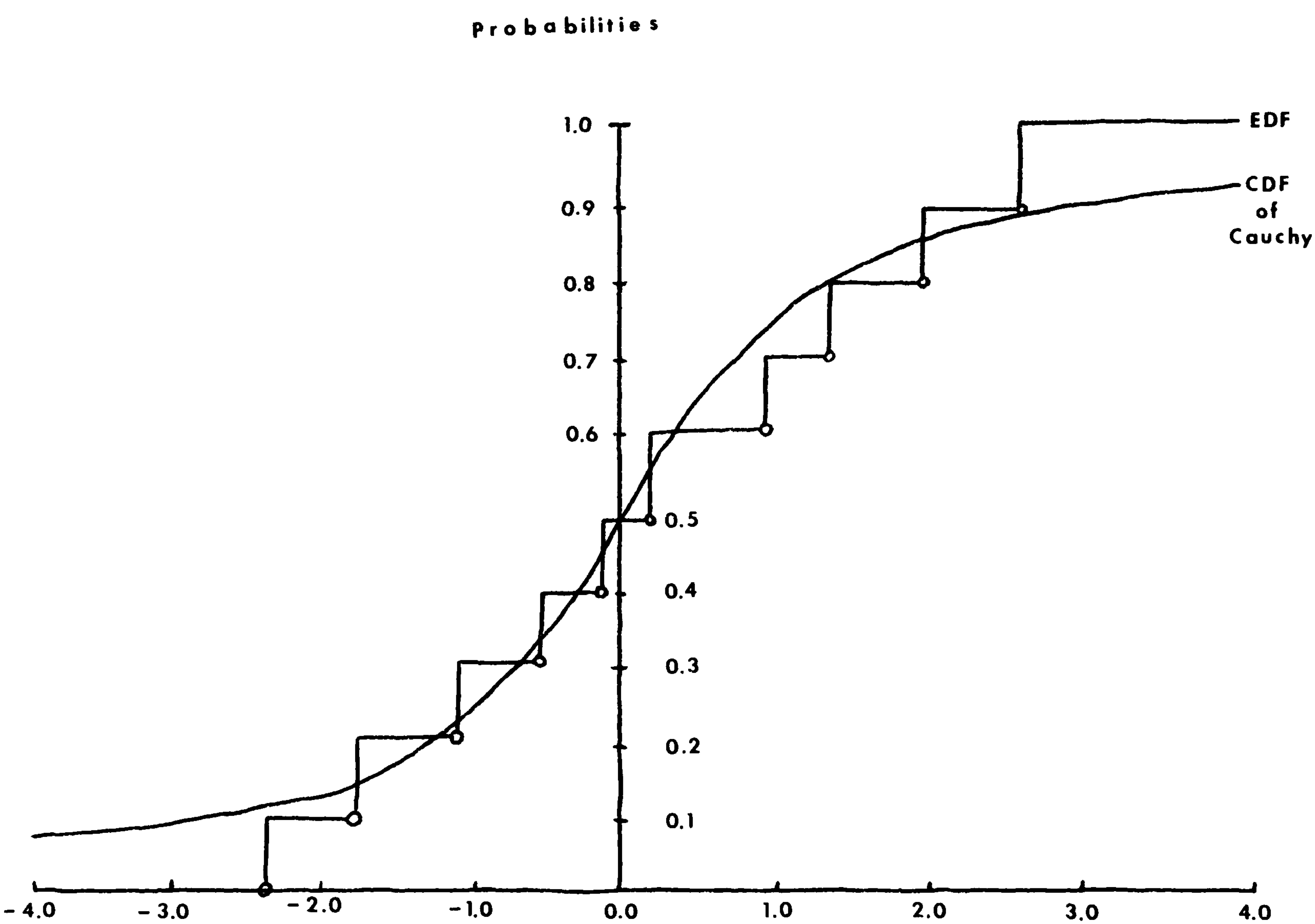


Figure 3. CDF versus EDF

value is retained along with the parameter value that gave the minimum discrepency. This procedure is repeated until the minimum value over a wide range of parameter values is obtained.

Minimum distance estimation is closely related in theory to another statistical procedure called goodness-of-fit. In a goodness-of-fit test, one tries to determine if a particular set of data belongs to a particular probability distribution. The sample data is used to determine the edf and to estimate the parameters of the theoretical distribution. The discrepency between the edf and the theoretical distribution is then calculated and compared to a value in a goodness-of-fit table to determine the data's confidence level in belonging to the theoretical distribution. MD estimation extends the goodness-of-fit process by minimizing the discrepency through alteration of the distribution's parameters and thereby estimating those parameters.

The descrepency in MD estimation or a goodness-of-fit test can be defined in a number of ways because the methods do not depend upon any particular defintion of distance (29:204). In fact, several distance defintions are discussed in this chapter and employed in this thesis.

Weighted Kolmogorov Distance

The weighted Kolmogorov distance is defined as

$$D_{\phi}(Gn,F_{\theta}) = \sup_{x} | Gn(x) - F_{\theta}(x) | \Phi(F_{\theta}(x)). \qquad 4\text{-}2$$

This distance measure is well known from its use in the Kolmogorov-Smirnov goodness-of-fit tests. The distance is defined as the maximum vertical distance (also known as sup) between the edf evaluated at a given point and the value of the theoretical distribution at the same point. The weighting function, $\Phi(F_\theta(x))$, assigns weights for each point and in this thesis is chosen to be equal to one.

When using censored samples, the distances between censored variates and the edf are not used in determining the Kolmogorov distance. Everything else remains the same. The value of the edf does not change and still uses the original sample size and variate postion in determining its value.

<u>Weighted Cramer-von Mises (CVM) Distance</u>

The weighted CVM distance is defined mathematically as

$$W_\Phi^2(Gn,F_\theta) = \int_{-\infty}^{\infty} (Gn(x) - F_\theta(x))^2 \Phi(F_\theta(x)) dF_\theta(x). \qquad 4\text{-}3$$

This distance represents the area discrepency between the edf and the theoretical distribution. As with the Kolmogorov distance, there is a weighting function, $\Phi(F_\theta(x))$, which is assigned the value of one.

For censored samples, the distance measure definition supplied by Sweeder (28:208) may be used which is

$$D_\Phi^2(Gn,F_\theta) = \int_{x_{min}}^{x_{max}} (Gn(x) - F_\theta(x))^2 \Phi(F_\theta(x)) dF_\theta(x). \qquad 4\text{-}4$$

For the CVM discrepency with ϕ equal to one, this is defined as

$$W^2 = \sum_{i=3}^{n-3} \int_{x_i}^{x_{i+1}} (i/n - F_\theta(x))^2 dF_\theta(x). \qquad 4\text{-}5$$

This can be rewritten as

$$\sum_{i=3}^{n-3} \int_{x_i}^{x_{i+1}} ((i/n)^2 - (i/n)F_\theta(x) + F_\theta(x)^2) dF_\theta(x) \qquad 4\text{-}6$$

or

$$\sum_{i=3}^{n-3} \left[\int_{x_i}^{x_{i+1}} (i/n)^2 dF_\theta(x) - \int_{x_i}^{x_{i+1}} (i/n) F_\theta(x) dF_\theta(x) + \int_{x_i}^{x_{i+1}} F_\theta(x)^2 dF_\theta(x) \right] . \qquad 4\text{-}7$$

Integrating yields

$$\sum_{i=3}^{n-3} \left[\left((i/n)^2 F_\theta(x) - (i/(2n)) F_\theta^2(x) + (1/3) F_\theta^3(x) \right) \Bigg|_{x_i}^{x_{i+1}} \right] \qquad 4\text{-}8$$

which is the Cramer-von Mises distance for censored samples.

Anderson-Darling (AD) Statistic

This distance is represented mathematically as

$$A_\phi^2(Gn, F_\theta) = \int_{-\infty}^{\infty} (Gn(x) - F_\theta(x))^2 (1/(u(1-u))) dF_\theta(x). \qquad 4\text{-}9$$

This distance is actually the Cramer-von Mises statistic with the weighting function equal to $(1/(u(1-u)))$ and u

equal to the theoretical distribution. It can be seen by inspection that this weighting function gives more weight to the tails of the distribution.

For censored samples, this measure is defined as

$$A^2 = \sum_{i=3}^{n-3}\left[\int_{x_i}^{x_{i+1}} (i/n - F_\theta(x))^2 (1/(F_\theta(x)(1-F_\theta(x))))dF_\theta(x)\right]. \quad 4\text{-}10$$

Rewriting and integrating yields

$$\sum_{i=3}^{n-3}\left[\left((i/n)^2 \ln(F_\theta(x)/(1-F_\theta(x))) + 2(i/n)\ln(1-F_\theta(x)) - F_\theta(x) - \ln(1-F_\theta(x))\right)\Bigg|_{x_i}^{x_{i+1}}\right]. \quad 4\text{-}11$$

The Kuiper Distance

The Kuiper distance is defined as

$$V(Gn,F_\theta) = \sup_{-\infty<a<b<\infty} |(Gn(b) - Gn(a)) - (F_\theta(b) - F_\theta(a))|. \quad 4\text{-}12$$

This distance is related to the Kolmogorov distance in that it also uses the vertical distance between the edf and the theoretical distribution. Two distances are calculated, however. The first distance is from the edf to the theoretical distribution where the edf value is greater than the theoretical. The second distance is from the theoretical distribution to the edf where the theoretical value is greater than the edf. These distances are illustrated in

Figure 4. Sup values are recorded for each, summed, and the sum is then minimized using MD techniques. The points a and b represent the points at which the sups occur and define the maximal probability interval. This distance tends to give more accurate estimates for scale parameters than it does for location parameters (18:38). Censored samples are handled in the same way as the Kolmogorov distance.

<u>Watson Distance</u>

The Watson distance is derived from a general class of discrepencies defined as

$$Z^2_{a,b} \quad (Gn, F_\theta) \;=\; a \int_{-\infty}^{\infty} ((Gn(x) - F_\theta(x))^2 \, dF_\theta(x)$$

$$+ \; b\left[\int_{-\infty}^{\infty} (Gn(x) - F_\theta(x))dF_\theta(x)\right]^2 . \qquad 4\text{-}13$$

When the parameter a and b equal 1 and 0 respectively, the discrepency reduces to the Cramer-von Mises statistic. When a equals 1 and b equals -1, the discrepency is defined as the Watson statistic (24:616). As with the Kuiper distance, the Watson distance tends to give better estimates of the scale parmeter (24:618).

The censored sample distance is defined as

$$Z^2 \;=\; W^2 - \left[\sum_{i=3}^{n-3} \int_{x_i}^{x_{i+1}} (i/n - F_\theta(x))dF_\theta(x)\right]^2 . \qquad 4\text{-}14$$

Integrating, this yields

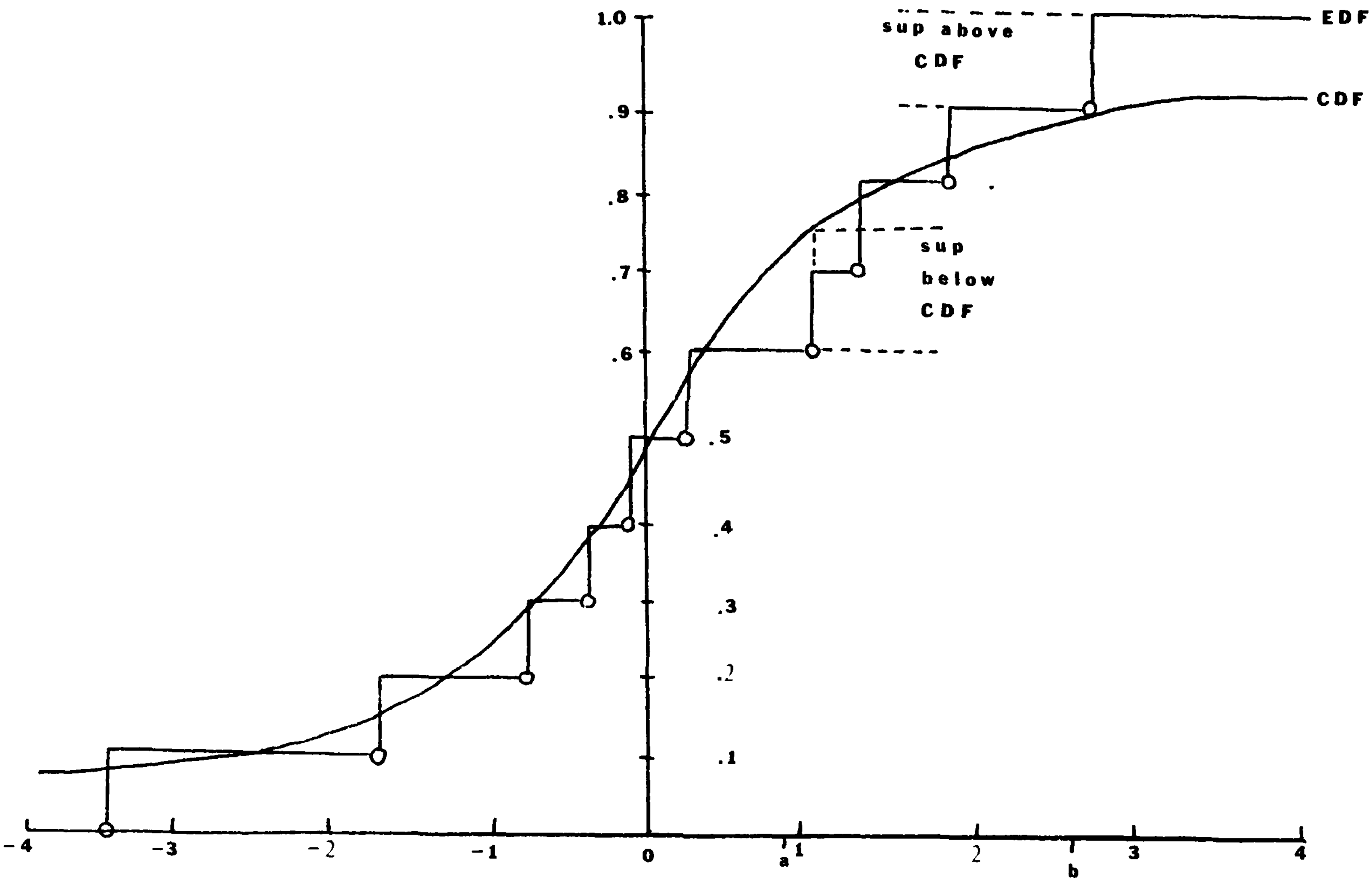


Figure 4. Kuiper Distance

$$W^2 - \left[\sum_{i=3}^{n-3}\left((i/n)F_\theta(x) - (1/2)F_\theta(x)\right)\Big|_{x_i}^{x_{i+1}}\right]^2 \quad 4\text{-}15$$

which is the Watson distance measure for censored samples.

Empirical Distribution Function Distance

Another class of distances is based upon the empirical characteristic function (25:1206). Using equation 2-5 as the empirical characteristic function and letting

$$C(t;\theta) = \int_{-\infty}^{\infty} e^{itz}\, dGn(z), \quad -\infty < t < +\infty \quad , \qquad 4\text{-}16$$

typical distance measures include

$$\delta_p(Gn,F_\theta) = \int_{-\infty}^{\infty} |Cn(t) - C(t;\theta)|^p w(t)dt \qquad 4\text{-}17$$

and

$$\delta(Gn,F_\theta) = \sup_t |\, Cn(t) - C(t;\theta)|\, w(t)^{1/2} \ . \qquad 4\text{-}18$$

These distance measures were not investigated in this thesis but are recommended as possible follow on work.

Summary

The Kolmogorov, Cramer-von Mises, Anderson-Darling, Kuiper, and Watson distance measures were used to estimate the location parameter of the Cauchy distribution. The Kuiper and Watson were included because the Parr study showed these estimators to perform better with samples from long tailed

distributions (24:620). In doing the location estimation, the value used for the scale parameter was the mle estimate of the scale and that value was retained throughout the minimum distance estimation. The same process was repeated in estimating the scale parameter. The Kuiper and Watson measures were used to estimate the scale parameter and the mle estimate of the location was used and retained throughout the MD process. Results of the minimum distance estimations are discussed in Chapter Six.

V. Monte Carlo Analysis

Introduction

Monte Carlo analysis is the process which is used to evaluate the efficiency of the estimators. The process involves generation of the Cauchy deviates, computation of the parameter estimates, comparison of the estimates with the real parameter values, and finally comparison of the estimators with each other. All of the Monte Carlo analysis in this thesis was done with the Vax 11/780 computer which is owned and operated by AFIT.

Generation of Data

The generation of Cauchy deviates was facilitated by the use of the IMSL subroutine GGCAY. This routine was also used by Koutrouvelis in his estimations using the empirical characteristic function. GGCAY uses a ratio method to generate the deviates as opposed to the inversion method because of the ratio's better efficiency (15:GGCAY-1). Standard deviates were generated which had the scale value of one and the location value of zero. Deviates from Cauchy distributions other than the standard were generated by multiplying the standard deviate by the scale value and adding to it the location value. Sample sizes of 6,8,10,12, and 16 were generated with location and scale values of zero and one respectively. Sample sizes of 6 and 16 were generated with a location value of -2 and a scale value of 5 to test for invariance. The number of samples generated for each

sample size and location/scale set was 1000. Caso's study recommended 2000 samples but stated that 1000 was sufficient (5:37). The sample size of 1000 was used to reduce the load on computer resources.

Initial Estimates

The initial estimates were the median for the location and the modified semi-interquartile range for the scale. Both estimates were computed in subroutine MEDSEM which was adapted from the subroutine used to generate the maximum likelihood estimators in the Princeton Study. A prerequisite to running the routine was that the sample needed to be ordered.

Maximum Likelihood Estimates

The maximum likelihood estimators were generated using a subroutine called CMLE which was also adapted from the Princeton study (1:262). This routine required the same Cauchy ordered sample and the median and modified semi-interquartile range generated from MEDSEM. The routine then used an iterative procedure to generate the mles. One drawback to numerical estimates of the mles was the possibility of divergence or non-convergence. This routine as initially written provided for only a maximum of twenty iterations for an estimate to converge, and hence produced for a number of non-converging estimates, especially in the smaller sample sizes. Results of non-converging estimates are shown in Table I in Appendix A. All of the scale estimates converged

and none of the non-converging location estimates diverged. As shown in Table I, an iteration maximum of 100 produced only one non-converging estimate in a sample size of six and no non-converging estimates for larger sample sizes. The difference between the final nonconverging mle estimates was less than .004. Because of these results, the iteration maximum of 100 was used in CMLE.

Basically, the iteration scheme proceeded as follows. First, the median and semi-interquartile range was put into Equations 3-5 to estimate the location and 3-6 to estimate the scale. The differences between the new estimates and the old estimates were computed. If the absolute difference between the new estimate and old estimate of location was less than .001 times the scale estimate, the estimate converged. If the absolute difference between the new estimate and the old estimate of scale was less than .05 times the scale estimate, the scale estimate converged. These were the criteria for convergence used in the Princeton Study. If both estimates converged, the new estimates were returned as the mle estimates of location and scale. If one or both estimates did not converge, the iteration process repeated until convergence occured or until the iteration maximum was reached. If the iteration maximum was reached, the final estimates were used as the mle estimates.

Minimum Distance Estimators

The minimum distance estimators and ways to estimate

them were discussed in Chapter Four. Only one parameter was estimated at a time. The other parameter was held fixed at its mle estimate while the other one was estimated by MD methods. Let λ' represent the mle estimate of location and ψ' represent the mle estimate of scale. To find which location estimate gave the minimum distance, location estimates in the range from $\lambda' \pm \psi'$ were used to calculate the distance. Starting with a location estimate equal to λ', the estimate was incremented by $.01\psi'$ until it reached a value value of $\lambda'+\psi'$. If the minimum distance occured at $\lambda'+\psi'$, then the estimate was extended to $\lambda'+2\psi'$, and so forth. The estimate that yielded the minimum distance was recorded. The process was repeated again starting with λ' and $.01\psi'$ being subtracted from each previous estimate till the estimate reached a value of $\lambda'-\psi'$. As before, if the minimum distance occurred at $\lambda'-\psi'$, the process was repeated to $\lambda'-2\psi'$, etc. Finally, the estimate that yielded the minimum distance was used as the estimate for the sample. The scale parameter was estimated in the same same way except that the estimator range was from to $\psi' \pm .5\psi'$ to prevent the scale estimate from going below zero.

The distance measures for non-censored samples was computed using the formulas provided by Stephens (27:731). The only requirement for use of these formulas was that F(x) be continuous and completely specified; hense these formulas can be used with the Cauchy distribution. Formulas for censored samples were derived in Chapter Four.

Kolmogorov Distance Estimator. The Kolmogorov minimum distance estimator for non-censored samples is defined by Stephens as (27:731):

$$D^{+} = \max(1 \leq i \leq n) \quad [(i/n) - z_i]$$
$$D^{-} = \max(1 \leq i \leq n) \quad [z_i - (i/n)]$$
$$D = \max(D^{+}, D^{-}) \qquad 5\text{-}1$$

where z_i , in this thesis, is a variate from the Cauchy cdf. The minimum distance is calculated in the subroutine KSMIN. KSMIN first calculates sup using the mle estimates and then calls on the subroutine VEDIS. VEDIS calculates both D^{+} and D^{-}. When these values are returned to KSMIN, D is determined This distance is stored and the procedure described in the previous section is initiated and proceeds until the estimate with the minimum distance is determined. Computation for censored samples was discussed in Chapter Four. The only difference is that i is restricted to be between 3 and n-2. This estimator was used as a location estimator only.

Cramer-von Mises Estimator. This distance estimator was also used only as a location estimator. The computational formula for non-censored samples is defined by Stephens as (27:731):

$$W^2 = \sum_{i=1}^{n} [z_i - (2i-1)/2n] + (1/12n). \qquad 5\text{-}2$$

The computation of the minimum distance is done in the subroutine AEMIN. AEMIN is a general subroutine that computes the minimum distance for the CVM, the AD, and the Watson

(location) estimators. AEMIN calls upon AEDIS to calculate the distance. AEDIS calculates the distance in accordance with the value of IFLAG where a value of 0 calculates the CVM distance, a value of 1 calculates the AD distance, and a value of 2 calculates the Watson distance. The user supplies the value of IFLAG in the main program. The minimum distance calculated in AEMIN is also computed as described in the MD estimator section. For censored samples, AEDIS is modified to use equation 4-8.

Anderson-Darling Estimator. The computational formula for the AD estimator is defined as Stephens as (27:731):

$$A^2 = - \left(\sum_{i=1}^{n} (2i-1)[\ln z_i + \ln(1-z_{n+1-i})] \right)/n - n \qquad 5\text{-}3$$

The minimum distance using this distance is also done by AEMIN as described in the previous section. This distance measure is used for location estimates only. The computational formula for censored samples is given by 4-11. AEDIS is modified to compute the distance for censored samples as well.

Kuiper Estimator. Stephens presents the computational formula for the Kuiper distance as (27:731):

$$V = D^+ + D^- \qquad 5\text{-}4$$

The subroutine that computes the minimum distance is KUMIN. KUMIN calls on VEDIS to calculate D^+ and D^- . KUMIN then adds these two distances and finds the minimum distance as described earlier. Both the location and scale parameter

are estimated using this distance. For censored samples, the procedure is similar to finding the Kolmogorov distance for censored samples.

Watson Estimator. The computational formula for the Watson distance using non-censored samples is defined by Stephens as (27:731):

$$U^2 = W^2 - n(\tilde{z} - .5) \qquad 5\text{-}5$$

where

$$\tilde{z} = \sum_{i=1}^{n} z_i / n \qquad 5\text{-}6$$

For estimating the location parameter, AEMIN calculates the minimum distance and calls upon AEDIS to calculate the distance for a particular location estimate. To estimate the scale parameter, the subroutine WAMIN calculates the minimum distance and calls upon AEDIS to calculate the distance for a particular estimate of scale. For censored samples, equation 4-15 is used to estimate the location parameter. This equation did not give good estimates for the scale parameter when using censored samples.

Evaluation Criteria

The efficiency or accuracy of an estimator can be measured in several ways. A common method is to calculate how close the estimate actually comes to the parameter being estimated. This difference is squared, summed over the number of samples, and divided by the number of samples. This measure is defined as the mean squared error and is repre-

sented mathematically as

$$MSE = \sum_{i=1}^{n} (\theta - \theta_i')^2 / n \qquad 5\text{-}7$$

where θ is the parameter and θ' is the parameter estimate. The MSE for each estimate including the mle estimates, the median, the modified semi-interquartile range, and the minimum distance estimates are recorded. One problem with the MSE is that it is not scale invariant (18:58). Hense, if one wants to test the invariance of the estimator, another measure should be used.

Another measure called the relative efficiency (RE) was found to be used extensively throughout the literature. This measure compares the MSE of one estimator to the MSE of another and is defined mathematically as

$$RE = MSE_k / MSE_i \qquad 5\text{-}8$$

In this thesis MSE_i will represent the MSE of the mle estimators and MSE_k will represent the MSE of the non-mle estimators. Thus an RE of less than one indicates an improvement over the mle and an RE of greater than one indicates no improvement over the mle. Since the RE is scale invariant, it will allow for investigation of the invariance of the MD estimators.

A Walk-Through the Computer Program

This section will briefly discuss how the computer program derived and evaluated the estimates. The main program that derived the location estimates is called Cauchy Loca-

tion Parameter Estimation and Efficiencies (CLPEE) and can be found in Appendix B. The program that derived the scale estimates is very similar to CPLEE and is called Cauchy Scale Parameter Estimation and Efficiencies (CSPEE). It can be found in Appendix C. The first step in either case was to generate the empirical distribution function since this function had to be computed only once. Next, a Cauchy random sample was generated using GGCAY and ordered from lowest to highest by a routine the main program. MEDSEM was then called upon to calculate the median and modified semi-iterquartile range. The differences between these values and the real parameters were calculated, squared, and added on to each previous squared difference. This process occurred after deriving each type of estimate. Next, CMLE was called to derive the mle estimates using the median and modified semi-interquartile range as initial estimates. After processing the MSEs, the subroutines KSMIN, AEMIN, and KUMIN were called to estimate the location parameter or KUMIN and WAMIN were called to estimate the scale parameter. After the final sample was processed, the REs were calculated and recorded. The subroutines VEDIS and AEDIS used to calculate the distances for non-censored samples can be found in Appendix B and C. The modified subroutines of VEDIS and AEDIS used to calculate the distances for censored samples can be found in Appendix D. A flow chart for CLPEE is shown in Figure 5. The flow of CSPEE is similar.

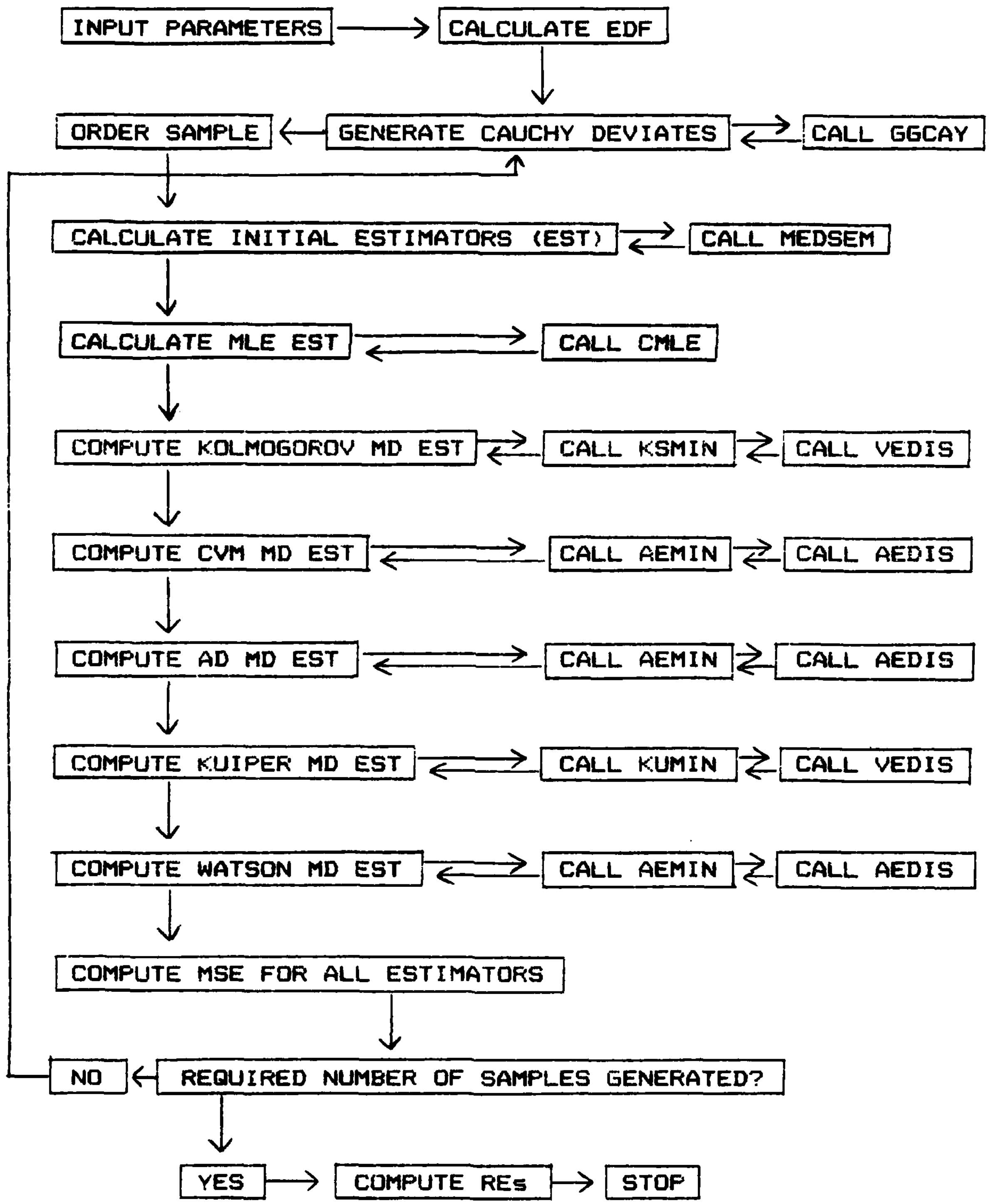


Figure 5. Flow Chart of Cauchy Location Parameter Estimation and Efficiencies (CLPEE)

VI. CONCLUSIONS AND RECOMMENDATIONS

Results

Tables II-1 through II-5 in Appendix A show the MSE and RE for the estimators of the (0,1) parameter set for non-censored samples. The tables show that the Kolmogorov, Cramer von-Mises, and Anderson Darling distance measures are inferior to the mle and inferior to the median as well. The Kuiper location measure was sometimes superior to the median and sometimes not. It was never as good as the mle. The Watson statistic was the one that appeared to be on the same level as the mle. For sample sizes 8 and 16 it actually outperformed the mle and in the other sample sizes it was very close. For the scale estimators, the results were similar. The Kuiper scale measure always outperformed the modified semi-interquartile range but was never as good as the mle. The Watson scale estimator was slightly better than the mle for all sample sizes except 6.

Because of closeness of the Watson location and scale estimators to the mle estimates, another run was made with 5000 samples and a new value for DSEED. The parameter values remained the same. The results are shown in Table III and in this case, the mle outperformed the Watson estimators, but only slightly. In all sample sizes, the Watson estimator was within 2% of the mle except for the location estimate with sample size 6.

Because the Kolmogorov, CVM, and AD estimators did not

perform as well as expected, it was thought that the reason might be because of the Cauchy's infinite mean of the first and last order statistic and the infinite variance of the second and next to last order statistic. For this reason, the same runs were made with the samples being censored as outlined in Chapter Four. The results are shown in Tables II-6 through II-10 in Appendix A. The location estimates for the three estimators did improve in the censored samples, but they were still not as good as the mle or the Kuiper and Watson estimators for non-censored samples. In most cases they were not as good as the median (In one case only, the Komogorov was slightly better than the median). The unusually high RE for the CVM distance in sample six was caused by one very high estimate. Similar REs were observed with different seeds and in sample sizes of seven as well. The problem cleared up when a sample size of at least eight was used. The Kuiper estimator became worse when censored samples were used and approached values very close to the median. The Watson estimator, although still good, was worse than its estimates for non-censored samples. For the scale estimates, the results using censored samples were worse. The Kuiper estimates were never as good as the modified semi-interquartile range and the Watson estimators failed to run at all. Since every Watson estimate was consistently bad, the MSEs and REs were not recorded.

Finally, the invariance of the estimators was tested. With the same seed values and using non-censored samples, sample sizes of 6 and 16 were run with a parameter set of

(-2,5). Table II-11 shows the results for a sample size of six and Table II-12 shows the results for a sample size of sixteen. A comparison of these tables with Tables II-1 and II-5 reveal that the RE for each estimator did not change significantly with the change in parameter values. In fact, the largest change was less than .1 per cent. Any change at all was due to round off error. Thus the minimum distance estimators for the Cauchy distribution are invariant to different paramteter values and thus Parr's assertion is confirmed (24:617).

Conclusions

The censored samples outperformed the non-censored samples in the case of the Kolmogorov, Cramer von-Mises, and Anderson-Darling estimators of location. However, the improvement was not enough to justify using them because their estimates were still not as good as the mle or the median. On the other hand, the non-censored samples outperformed the censored samples in the case of the Kuiper and Watson statistics. This was true for both the location and scale estimates. Clearly, the Watson statistic, using non-censored samples, was the best of all in either case as a location or scale estimator and was always almost as good as the mle. However, there was still no real improvement over the mle and hense no attempt was made to improve the opposite parameter, as described in the methodology section of Chapter One.

Since no real improvement over the mle estimate has

been found, what has been gained from this research? For one thing, this study did reveal that the Watson statistic is an excellent distance measure for the Cauchy distribution. Hense this distance measure might be useful in a goodness-of-fit test as described in Chapter Four. The goodness-of-fit test would be useful to determine if a particular data set is Cauchy distributed and the Watson distance measure might provide a more powerful test than now exists for the Cauchy. The Kuiper statistic might provide a powerful test as well.

Follow-on Work

The following topics are recommended as potential follow-on work:

1) Because the Kuiper and Watson estimators were very good estimators for the Cauchy distribution, they have a lot of potential in being used as distance measures in goodness-of-fit tests as previously described. One could use mle methods to estimate the parameters of the hypothesized Cauchy distribution and then use Monte Carlo analysis, with the Watson and Kupier statistics as distance measures, to derive the necessary goodness-of-fit tables. The power of the goodness-of-fit tests using these distance measures could be compared to other goodness-of-fit tests using the Kolmogorov, CVM, and AD distance measures.

2) There is an IMSL subroutine called ZXMIN which will minimize a function of several variables. Parr used this

routine to determine the minimum distance estimators in his minimum distance study (24:621). One could use this routine to estimate both parameters simultaneously using the median and semi-interquartile range as initial estimates. A comparison between the mle and Watson estimators could then be made to measure the efficiency in terms of the computer time required to calculate the estimates. The accuaracy of the estimators could be double-checked as well. Also, the robustness properties of the MD estimators could be explored and compared with the robustness properties of the mle. This would be done by generating samples from different distributions, such as the normal, and using these samples to estimate the parameters. Even if the Watson estimators are not as accurate as the mle, perhaps the speed at which they can be calculated and their robustness properties might justify their use.

3) Finally, there is another distance estimator derived from the characteristic function, which was described in Chapter Four. This estimator could also be used to estimate the Cauchy parameters and its efficiency compared to other estimators. An estimator to compare with the characteristic function distance measure might be the one derived by Koutrouvelis (see Chapter Three) since he used the characteristic function to derive his estimator. This topic might be also expanded to other distributions that follow stable law and that do not have closed form expressions such as the normal and Cauchy.

Bibliography

1. Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, W. H. Rogers, J.W. Tukey. Robust Estimates of Location, Survey and Advances. Princeton University Press, Princeton, New Jersey, 1972

2. Barnett, V.D. "Order Statistics Estimators of the Location of the Cauchy Distribution". American Statistical Association Journal, 61:1205-1218 (1966)

3. -------. "Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots". Biometrika, 53:151-164 (1966)

4. Bertrand, 2Lt David E. Comparison of Estimation Techniques for the Four Parameter Beta Distribution. MS Thesis, AFIT/GOR/MA/81D-1. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, December 1981

5. Caso, J. Robust Estimation Techniques for Location Parameter Estimation of Symetric Distributions, MS Thesis. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, 1972.

6. Chamberlain, 1Lt Robert B. Nearly Best Linear Unbiased Estimators of the Location and Scale Parameters of the Cauchy Distribution by Use of Censored Order Statistics. MS Thesis. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH. 1968

7. Conover, W.J. Practical Nonparametric Statistics. New York: John Wiley Sons, Inc., 1971

8. Copas, J. B. " On the unimodality of the likelihood for the Cauchy distribution ". Biometrika, 62: 701-704 (1975)

9. Derman, Cyrus, Leon J. Gleser, and Ingram Olkin. A Guide to Probability Theory and Application. Holt, Rinehart, and Winston, Inc., 1973

10 Deutsch, Ralph. Estimation Theory. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1965

11. Granger, W. J., and Daniel Orr. ""Infinite Variance" and Research Strategy in Time Series Analysis". Journal of the American Statistical Association, 67: 275-285 (1972)

12. Haas, Gerald, Lee Bain, and Charles Antle. " Inferences for the Cauchy distribution based on maximum likelihood estimators". Biometrika, 57: 403-408 (1970)

13. Higgins, J.J., and D.M.Tichenor. "Window Estimates of Location and Scale With Applications to the Cauchy Distribution". Applied Mathematics and Computation, 3: 113-126 (1977)

14. -------. "Efficiencies for Window Estimates of the Parameters of the Cauchy Distribution". Applied Mathematics and Computation, 4: 157-165 (1978)

15. International Mathematical and Statistical Library Reference Manual-0008. Houston: IMSL, 1980

16. Johnson, N.L., and S. Kotz. Continuous Univariate Distributions, Volume I. Boston: Houghton Mifflin Co., 1970.

17. Jonson, Capt Edward C. Conditional Nearly Best Linear Estimation of the Location and Scale Parameters of the Cauchy Distribution by the Use of Censored Order Statistics. MS Thesis, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH. 1969

18. Keffer, 2Lt Jim H. Robust Minimum Distance Estimation of the Three Parameter Lognormal Distribution. MS Thesis, AFIT/GOR/MA/83D-3. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH. 1983

19. Kendall, Maurice G., ScD. and William R. Buckland, B.Sc., PhD. A Dictionary of Statistical Terms. Hafner Publishing Co., New York, 1960

20. Koutrouvelis, Ioannis. "Regression-Type Estimation of the Parameters of Stable Laws". Journal of the American Statistical Association, 75:918-928 (1980)

21. -------. "Estimation of location and scale in Cauchy distributions using the empirical characteristic function". Biometrika, 69:205-213 (1982)

22. Lennik, Yu V. The Method of Least Squares. Pergamon Press, 1961

23. Moore, Lieutenant Harry Richard II, United States Navy. Robust Regression Using Maximum-Likelihood Weighting And Assuming Cauchy-Distributed Random Error. MS Thesis, Naval Posrgraduate School, Monterey, California. 1977(AD-A045132)

24. Parr, William C. and William R. Schucany. "Minimum Distance and Robust Estimation". Journal of the American Statistical Association, 75: 616-624 (1980)

25. Parr, William C. "Minimum Distance Estimation: A Bibligraphy". Communication Statistics-Theoritical Methods, A10(12), 1205-1224 (1981)

26. Robinson, David W. and Peter A. W. Lewis. Generating Gamma and Cauchy Random Variables: An Extension to the Naval Postgraduate School Random Number Package. Technical Report #NPS-72Ro75041. Naval Postgraduate School, Monterey, California. 1975(AD-A057230)

27 Stephens, M. A.. "EDF Statistics for Goodness of Fit and Some Comparisons". Journal of the American Statistical Association, 69: 730-737 (1974)

28. Sweeder, Capt James. Nonparametric Estimation of Distribution and Density Functions with Applications. Dissertation, AFIT/DS/MA/82-1. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH. 1983

29. Wolfowitz,J. "Estimation by the minimum distance method". Annals of the Institute of Statistical Mathematics, 5: 9-23 (1953)

30. -------. "Estimation by the minimum distance method in nonparametric stochastic difference equations". Annals of Mathematical Statistics, 25:203-217 (1954)

31. -------. "The Minimum Distance Method". Annals of Mathematical Statistcs, 28: 75-88 (1957)

32. Woodroofe, Michael. Probability with Applications. McGraw-Hill Book Company, New York, 1975

Appendix A

Tables

TABLE I

Non-Converging MLE Estimates (1000 Samples)

Iteration Maximum of 20

Sample Size	Non-Converging Estimates
6	75
8	28
10	5
12	2
16	0

Iteration Maximum of 50

Sample Size	Non-Converging Estimates
6	7
8	2
10	0
12	0
16	0

Iteration Maximum of 100

Sample Size	Non-Converging Estimates
6	1
8	0
10	0
12	0
16	0

TABLE II-1

MSE and RE of Estimators

Sample Size: 6 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Non-censored Samples

Location Estimator	MSE	RE
Median	.8157362	1.275193
Maximum Likelihood	.6396965	-
Kolmogorov	1.0176807	1.590879
Cramer von-Mises	.9793309	1.530931
Anderson-Darling	.9884748	1.545225
Kuiper	.6723628	1.051065
Watson	.6968500	1.089345

Scale Estimator	MSE	RE
Mod Semi-Interquartile Range	.8787380	1.577240
Maximum Likelihood	.5571364	-
Kuiper	.5901666	1.059286
Watson	.5886748	1.056608

TABLE II-2

MSE and RE of Estimators

Sample Size: 8　　DSEED: 4385673
Location Parameter: 0　　Scale Parameter: 1
Number of Samples: 1000

Non-censored Samples

Location Estimator	MSE	RE
Median	.5876736	1.099047
Maximum Likelihood	.5347122	-
Kolmogorov	.7218777	1.350031
Cramer von-Mises	.6334487	1.184654
Anderson-Darling	.6576732	1.229957
Kuiper	.5523942	1.033068
Watson	.4949655	.925667
Scale Estimator	**MSE**	**RE**
Mod Semi-Interquartile Range	.6879840	1.504491
Maximum Likelihood	.4572869	-
Kuiper	.4698043	1.027373
Watson	.4529167	.990443

TABLE II-3

MSE and RE of Estimators

Sample Size: 10 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Non-censored Samples

Location Estimator	MSE	RE
Median	.3488217	1.132660
Maximum Likelihood	.3079669	-
Kolmogorov	.5114189	1.660629
Cramer von-Mises	.3902656	1.267232
Anderson-Darling	.4125222	1.339502
Kuiper	.3627019	1.177730
Watson	.3081932	1.000735
Scale Estimator	**MSE**	**RE**
Mod Semi-Interquartile Range	.4591780	1.605180
Maximum Likelihood	.2860602	-
Kuiper	.3225700	1.127630
Watson	.2656335	.928593

TABLE II-4

MSE and RE of Estimators

Sample Size: 12 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Non-censored Samples

Location Estimator	MSE	RE
Median	.2942365	1.223874
Maximum Likelihood	.2404139	-
Kolmogorov	.3974717	1.653281
Cramer von-Mises	.3191490	1.327498
Anderson-Darling	.3374594	1.403660
Kuiper	.2810406	1.168986
Watson	.2441652	1.015603
Scale Estimator	**MSE**	**RE**
Mod Semi-Interquartile Range	.2924091	1.328926
Maximum Likelihood	.2200342	-
Kuiper	.2485952	1.129802
Watson	.2182684	.991975

TABLE II-5

MSE and RE of Estimators

Sample Size: 16 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Non-censored Samples

Location Estimator	MSE	RE
Median	.1928420	1.152691
Maximum Likelihood	.1672972	-
Kolmogorov	.2593825	1.550429
Cramer von-Mises	.2113738	1.263462
Anderson-Darling	.2240063	1.338972
Kuiper	.2056511	1.229256
Watson	.1665640	.995617
Scale Estimator	**MSE**	**RE**
Mod Semi-Interquartile Range	.2096114	1.330707
Maximum Likelihood	.1575189	-
Kuiper	.1663357	1.055973
Watson	.1563607	.992648

TABLE II-6

MSE and RE of Estimators

Sample Size: 6 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Censored Samples

Location Estimator	MSE	RE
Median	.8157362	1.275193
Maximum Likelihood	.6396945	-
Kolmogorov	.8151365	1.274255
Cramer von-Mises	32.71764	51.14557
Anderson-Darling	.8151777	1.274320
Kuiper	.8085236	1.263917
Watson	.7480507	1.169384

Scale Estimator	MSE	RE
Mod Semi-Interquartile Range	.8787378	1.577240
Maximum Likelihood	.5571364	-
Kuiper	.2653000	4.761850
Watson	-	-

TABLE II-7

MSE and RE of Estimators

Sample Size: 8 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Censored Samples

Location Estimator	MSE	RE
Median	.5876736	1.099047
Maximum Likelihood	.5347122	-
Kolmogorov	.6621354	1.238303
Cramer von-Mises	.5948694	1.112504
Anderson-Darling	.6033888	1.256496
Kuiper	.6718636	1.256496
Watson	.5576661	.042928

Scale Estimator	MSE	RE
Mod Semi-Interquartile Range	.6879840	1.504491
Maximum Likelihood	.4572869	-
Kuiper	1.270862	2.779135
Watson	-	-

TABLE II-8

MSE and RE of Estimators

Sample Size: 10 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Censored Samples

Location Estimator	MSE	RE
Median	.3488217	1.132660
Maximum Likelihood	.3079669	-
Kolmogorov	.4796595	1.557503
Cramer von-Mises	.3603502	1.170094
Anderson-Darling	.3715640	1.206506
Kuiper	.3900580	1.266448
Watson	.3348452	1.087276
Scale Estimator	**MSE**	**RE**
Mod Semi-Interquartile Range	.4591780	1.605180
Maximum Likelihood	.2860602	-
Kuiper	.5427296	1.897257
Watson	-	-

TABLE II-9

MSE and RE of Estimators

Sample Size: 12 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Censored Samples

Location Estimator	MSE	RE
Median	.2942365	1.223874
Maximum Likelihood	.2404139	-
Kolmogorov	.3703235	1.540358
Cramer von-Mises	.2957745	1.230272
Anderson-Darling	.3054909	1.270687
Kuiper	.2920988	1.214983
Watson	.2707492	1.126179

Scale Estimator	MSE	RE
Mod Semi-Interquartile Range	.2924091	1.328926
Maximum Likelihood	.2200342	-
Kuiper	.3248410	1.476320
Watson	-	-

TABLE II-10

MSE and RE of Estimators

Sample Size: 16 DSEED: 4385673
Location Parameter: 0 Scale Parameter: 1
Number of Samples: 1000

Censored Samples

Location Estimator	MSE	RE
Median	.1928420	1.152691
Maximum Likelihood	.1672972	-
Kolmogrov	.2468639	1.475600
Cramer von-Mises	.2012539	1.202972
Anderson-Darling	.2089179	1.248783
Kuiper	.1998791	1.194754
Watson	.1795871	1.073461

Scale Estimator	MSE	RE
Mod Semi-Interquartile Range	.2096114	1.330707
Maximum Likelihood	.1575189	-
Kuiper	.1906907	1.210590
Watson	-	-

TABLE II-11

MSE and RE of Estimators

Sample Size: 6 DSEED: 4385673
Location Parameter: -2 Scale Parameter: 5
Number of Samples: 1000

Non-censored Samples

Location Estimator	MSE	RE
Median	20393.40	1.275192
Maximum Likelihood	15992.41	-
Kolmogorov	25442.00	1.590879
Cramer von-Mises	24483.27	1.530931
Anderson-Darling	24710.00	1.545108
Kuiper	16809.07	1.051066
Watson	17421.26	1.089345

Scale Estimator	MSE	RE
Mod Semi-Interquartile Range	21968.45	1.577240
Maximum Likelihood	13928.41	-
Kuiper	14754.24	1.059291
Watson	14727.70	1.057386

TABLE II-12

MSE and RE of Estimators

Sample Size: 16 DSEED: 4385673
Location Parameter: -2 Scale Parameter: 5
Number of Samples: 1000

Non-censored Samples

Location Estimator	MSE	RE
Median	4821.046	1.152689
Maximum Likelihood	4182.435	-
Kolmogorov	6484.562	1.550427
Cramer von-Mises	5284.348	1.263462
Anderson-Darling	5600.435	1.339037
Kuiper	5141.279	1.229255
Watson	4164.103	.995617
Scale Estimator	**MSE**	**RE**
Mod Semi-Interquartile Range	5240.282	1.330707
Maximum Likelihood	3937.969	-
Kuiper	4158.393	1.055974
Watson	3909.022	.992649

TABLE III

MSE and RE of Watson Estimator

Location Parameter: 0   Scale Parameter: 1
DSEED: 39927541
Number of Samples: 5000

Non-censored Samples

Location Estimator	MSE	RE
Sample Size		
6	.7371190	1.147544
8	.3979722	1.017078
10	.2891362	1.018591
12	.2228722	1.010221
16	.1566628	1.002658
Scale Estimator	**MSE**	**RE**
6	.6068876	1.007577
8	.4118958	1.010174
10	.2877366	1.003966
12	.2321404	1.003261
16	.1554233	1.002658

Appendix B

Computer Program Listing of CLPEE

```
C     ************************************************************
C     *                                                          *
C     *                      PROGRAM                             *
C     *                                                          *
C     *     CAUCHY LOCATION PARAMETER ESTIMATION AND             *
C     *     EFFICIENCIES (CLPEE)                                 *
C     *                                                          *
C     *     WRITTEN BY CAPT JOHN O SOURS, AFIT/GSO-85D, FOR      *
C     *     MS THESIS                                            *
C     *                                                          *
C     *                    DECEMBER 1985                         *
C     *                                                          *
C     ************************************************************
C
C
      PROGRAM  CLPEE
C
C     ************************************************************
C     *                                                          *
C     *                                                          *
C     *  PURPOSE: TO COMPARE PARAMETER ESTIMATION TECHNIQUES OF *
C     *           THE LOCATION PARAMETER OF THE TWO PARAMETER    *
C     *           CAUCHY DISTRIBUTION.  THE TECHNIQUES COMPARED *
C     *           ARE THE MAXIMUM LIKELIHOOD METHOD AND MINIMUM *
C     *           DISTANCE METHODS.                              *
C     *                                                          *
C     *                                                          *
C     * VARIABLES: R(20)    - ARRAY WHICH HOLDS THE CAUCHY       *
C     *                       DEVIATES                           *
C     *            EDF(20)- ARRAY WHICH HOLDS VALUES OF THE      *
C     *                       EMPIRICAL DISTRIBUTION FUNCTION    *
C     *            PI       - 3.1415926 RADIANS                  *
C     *            N        - SAMPLE SIZE                        *
C     *            RN       - REAL VALUE FOR SAMPLE SIZE         *
C     *            DSEED    - SEED VALUE USED FOR GENERATING     *
C     *                       CAUCHY DEVIATES IN GGCAY           *
C     *            WK(60)  - ARRAY USED IN GGCAY FOR WORKSPACE   *
C     *                       (3 TIMES R(I))                     *
C     *            LOC      - LOCATION PARAMETER VALUE OF THE    *
C     *                       CAUCHY DISTRIBUTION FROM WHICH THE *
C     *                       DEVIATES ARE TAKEN                 *
C     *            SCALE    - SCALE PARAMETER VALUE OF THE       *
C     *                       CAUCHY DISTRIBUTION FROM WHICH THE *
C     *                       DEVIATES ARE TAKEN                 *
C     *            IFLAG    - FLAG USED TO DETERMINE WHICH DIS-  *
C     *                       TANCE IS TO BE USED IN AEMIN       *
C     *                        (CRAMER-VON MISES,ANDERSON-DARLING,*
```

```
C     *                       OR WATSON)                          *
C     *            SSZE    - STEP SIZE USED IN THE MINIMUM          *
C     *                      DISTANCE TECHNIQUES                    *
C     *            SUMME   - SUM OF THE MEAN SQUARED ERRORS         *
C     *            SUMML     FOR THE MEDIAN, MLE LOCATION,          *
C     *            SUMKS     KOLMOGOROV DISTANCE,                   *
C     *            SUMCM     CRAMER VON-MISES DISTANCE,ANDERSON-    *
C     *            SUMAD     DARLING DISTANCE,KUIPER DISTANCE,      *
C     *            SUMKU     AND THE WATSON DISTANCE                *
C     *            SUMWA                                            *
C     *            EME     - SUMME/SUMML                            *
C     *            EKS     - SUMKS/SUMML                            *
C     *            ECM     - SUMCM/SUMML                            *
C     *            EAD     - SUMAD/SUMML                            *
C     *            EKU     - SUMKU/SUMML                            *
C     *            EWA     - SUMWA/SUMML                            *
C     *            DME     - DIFFERENCE BETWEEN THE LOC AND         *
C     *                      MEDIAN                                 *
C     *            DMEQ    - DME SQUARED (MEAN SQUARED ERROR)       *
C     *            DML     - DIFFERENCE BETWEEN THE LOC AND         *
C     *                      MLE OF LOCATION                        *
C     *            DMLQ    - DML SQUARED (MEAN SQUARED ERROR)       *
C     *            DKS     - DIFFERENCE BETWEEN THE LOC AND         *
C     *                      KOLMOGOROV ESTIMATE OF LOCATION        *
C     *            DKSQ    - DKS SQUARED (MEAN SQUARED ERROR)       *
C     *            DCM     - DIFFERENCE BETWEEN THE LOC AND         *
C     *                      CRAMER VON-MISES ESTIMATE OF           *
C     *                      LOCATION                               *
C     *            DCMQ    - DCM SQUARED (MEAN SQUARED ERROR)       *
C     *            DAD     - DIFFERENCE BETWEEN THE LOC AND         *
C     *                      ANDERSON-DARLING ESTIMATE OF           *
C     *                      LOCATION                               *
C     *            DADQ    - DAD SQUARED (MEAN SQUARED ERROR)       *
C     *            DKU     - DIFFERENCE BETWEEN THE LOC AND         *
C     *                      KUIPER ESTIMATE OF LOCATION            *
C     *            DKUQ    - DKU SQUARED (MEAN SQUARED ERROR)       *
C     *            DWA     - DIFFERENCE BETWEEN THE LOC AND         *
C     *                      WATSON ESTIMATE OF LOCATION            *
C     *            DWAQ    - DWA SQUARED (MEAN SQUARED ERROR)       *
C     *            MEDIAN  - THE MEDIAN OF THE SAMPLE               *
C     *            SEMIQ   - MODIFIED SEMI-INTERQUARTILE RANGE      *
C     *            MLEL    - MLE ESTIMATE OF LOCATION               *
C     *            MLES    - MLE ESTIMATE OF SCALE                  *
C     *            KSL     - KOLMOGOROV ESTIMATE OF LOCATION        *
C     *            CML     - CRAMER VON-MISES ESTIMATE OF           *
C     *                      LOCATION                               *
C     *            ADL     - ANDERSON-DARLING ESTIMATE OF           *
C     *                      LOCATION                               *
C     *            KUL     - KUIPER ESTIMATE OF LOCATION            *
C     *            WAL     - WATSON ESTIMATE OF LOCATION            *
C     *                                                             *
C     *                                                             *
C     ***************************************************************
```

```
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL WK(60),KSL,KUL,LOC,MLEL,MLES,MEDIAN
      DOUBLE PRECISION  DSEED
      PRINT*,'ENTER SAMPLE SIZE'
      READ*,N
      PRINT*,'ENTER # REPLICATIONS'
      READ*,NR
      PRINT*,'INPUT DSEED (0 FOR DEFAULT VALUE OF 4385673)'
      READ*,DSEED
      IF (DSEED.EQ.0) THEN
        DSEED=4385673.
      ENDIF
      PRINT*,'INPUT LOCATION AND SCALE PARAMETER'
      READ*,LOC,SCALE
      OPEN(UNIT=10,FILE='DATA',STATUS='NEW')
      WRITE(10,*)' '
      WRITE(10,*)'SAMPLE SIZE IS ',N,' LOC IS ',LOC,' SCALE IS
     )',SCALE
      WRITE(10,*)'  '
      PI=3.1415926
      SSZE=.01
      RN=N
      SUMME=0
      SUMML=0
      SUMKS=0
      SUMCM=0
      SUMAD=0
      SUMKU=0
      SUMWA=0
C
C     CALCULATE EMPIRICAL DISRIBUTION FUNCTION
C
      DO 6 I=1,N
         EDF(I)=I/RN
6     CONTINUE
      DO 75 L=1,NR
C
C     GENERATE CAUCHY DEVIATES
C
      CALL GGCAY (DSEED,N,WK,R)
      DO 5 I=1,N
           R(I)=R(I)*SCALE+LOC
5     CONTINUE
C
C     ORDER MATRIX OF CAUCHY DEVIATES
C
      DO 25 I=1,N
         DO 35 J=1,N-1
         IF (R(J).GT.R(J+1)) THEN
            TEMP=R(J)
```

```
              R(J)=R(J+1)
              R(J+1)=TEMP
           ENDIF
35         CONTINUE
25         CONTINUE
C
C     COMPUTE MEDIAN AND MODIFIED SEMI-INTERQUARTILE RANGE
C
      CALL MEDSEM (MEDIAN,SEMIQ)
      DME=LOC-MEDIAN
      DMEQ=DME**2
      SUMME=SUMME+DMEQ
C
C     CALCULATE MLE OF LOCATION AND SCALE
C
      CALL CMLE (MEDIAN,SEMIQ,MLEL,MLES)
C
      DML=LOC-MLEL
      DMLQ=DML**2
      SUMML=SUMML+DMLQ
C
C     MINIMIZE KOLMOGOROV DISTANCE TO ESTIMATE LOCATION PARAMETER
C
      CALL KSMIN (SSZE,MLEL,MLES,KSL)
      DKS=LOC-KSL
      DKSQ=DKS**2
      SUMKS=SUMKS+DKSQ
C
C     MINIMIZE CRAMER VON-MISES DISTANCE TO ESTIMATE LOCATION
C     PARAMETER
C
      IFLAG=0
      CALL AEMIN (SSZE,MLEL,MLES,CML,IFLAG)
      DCM=LOC-CML
      DCMQ=DCM**2
      SUMCM=SUMCM+DCMQ
C
C     MINIMIZE ANDERSON-DARLING DISTANCE TO ESTIMATE LOCATION
C     PARAMETER
C
      IFLAG=1
      CALL AEMIN (SSZE,MLEL,MLES,ADL,IFLAG)
      DAD=LOC-ADL
      DADQ=DAD**2
      SUMAD=SUMAD+DADQ
C
C     MINIMIZE KUIPER DISTANCE TO ESTIMATE LOCATION PARAMETER
C
      CALL KUMIN (SSZE,MLEL,MLES,KUL)
      DKU=LOC-KUL
      DKUQ=DKU**2
      SUMKU=SUMKU+DKUQ
```

```
C
C     MINIMIZE THE WATSON DISTANCE TO ESTIMATE THE LOCATION
C     PARAMETER
C
      IFLAG=2
      CALL AEMIN (SSZE,MLEL,MLES,WAL,IFLAG)
      DWA=LOC-WAL
      DWAQ=DWA**2
      SUMWA=SUMWA+DWAQ
75    CONTINUE
      EME=SUMME/SUMML
      EKS=SUMKS/SUMML
      ECM=SUMCM/SUMML
      EAD=SUMAD/SUMML
      EKU=SUMKU/SUMML
      EWA=SUMWA/SUMML
      WRITE(10,*)'SUMME = ',SUMME
      WRITE(10,*)'SUMML = ',SUMML
      WRITE(10,*)'SUMKS = ',SUMKS
      WRITE(10,*)'SUMCM = ',SUMCM
      WRITE(10,*)'SUMAD = ',SUMAD
      WRITE(10,*)'SUMKU = ',SUMKU
      WRITE(10,*)'SUMWA = ',SUMWA
      WRITE(10,*)' '
      WRITE(10,*)'EME = ',EME
      WRITE(10,*)'EKS = ',EKS
      WRITE(10,*)'ECM = ',ECM
      WRITE(10,*)'EAD = ',EAD
      WRITE(10,*)'EKU = ',EKU
      WRITE(10,*)'EWA = ',EWA
      STOP
      END
C
C
      SUBROUTINE MEDSEM (MEDIAN,SEMIQ)
C
C     ***********************************************************
C     *                                                         *
C     *  PURPOSE:  TO COMPUTE THE MEDIAN AND MODIFIED SEMI-INTER-*
C     *            QUARTILE RANGE OF THE CAUCHY RANDOM SAMPLE   *
C     *                                                         *
C     *                                                         *
C     * VARIABLES:  IH,IL,MH      -  INTEGER VARIABLES USED     *
C     *                              IN DETERMINING THE MODI-   *
C     *                              FIED SEMI-INTERQUARTILE    *
C     *                              RANGE                      *
C     *             D1,D2         -  DIFFERENCES BETWEEN THE    *
C     *                              MEDIAN AND CAUCHY RANDOM   *
C     *                              VARIATES                   *
C     *             SEMIQ1        -  VALUE USED IN AVERAGING    *
C     *                              THE RANGE IF THE SAMPLE    *
C     *                              SIZE IS EVEN AND EQUAL TO  *
```

```
C     *                                THE RANGE IF THE SAMPLE     *
C     *                                SIZE IS ODD                 *
C     *                                                            *
C     *  NOTE:  UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM *
C     *                                                            *
C     *                                                            *
C     **************************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL MEDIAN
      MH=(N+2)/2
      MEDIAN=.5*(R(N-MH+1)+R(MH))
      SEMIQ=0
      IH=N
      IL=1
      DO 1 I=1,MH
      D1 = MEDIAN-R(IL)
      D2 = R(IH)-MEDIAN
      SEMIQ1=SEMIQ
      IF (D1.GT.D2) GOTO 2
      IH=IH-1
      SEMIQ=D2
      GOTO 1
2     IL=IL+1
      SEMIQ=D1
1     CONTINUE
      IF (MOD(N,2).EQ.0) SEMIQ=.5*(SEMIQ+SEMIQ1)
      RETURN
      END
C
C
      SUBROUTINE CMLE (MEDIAN,SEMIQ,MLEL,MLES)
C
C
C     **************************************************************
C     *                                                            *
C     *  PURPOSE: TO CALCULATE THE MAXIMUM LIKELIHOOD ESTIMATORS*
C     *           OF THE LOCATION AND SCALE PARAMETERS FROM THE *
C     *           CAUCHY SAMPLE                                    *
C     *                                                            *
C     *                                                            *
C     * VARIABLES: IMAX   - MAXIMUM NUMBER OF ITERATIONS ALLOWED*
C     *            ITER   - ITERATION COUNTER                      *
C     *            MLELT  - VALUE OF MLE OF LOCATION/SCALE         *
C     *            MLEST    PARAMETERS AT THE BEGINNING OF EACH *
C     *                     ITERATION                              *
C     *            SUMO   - SUMMATION VARIABLES USED IN THE MLE *
C     *            SUM1     ITERATION PROCESS                      *
C     *            MLESSQ - MLES SQUARED                           *
C     *            Z      - VARIABLE USED IN THE MLE ITERATION    *
C     *                     PROCESS                                *
```

```
C     *               TMLES   - SQUARE ROOT OF THE MLE ESTIMATE OF  *
C     *                         SCALE                               *
C     *                                                             *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM      *
C     *                                                             *
C     *                                                             *
C     ***************************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL MLEL,MLES,MEDIAN,MLELT,MLEST,MLESSQ
      MLEL=MEDIAN
      MLES=SEMIQ
      IMAX=100
      ITER=0
20    MLELT=MLEL
      MLEST=MLES
      SUM0=0.
      SUM1=0.
      MLESSQ=MLES**2
      DO 10 I = 1,N
      Z=MLESSQ+(R(I)-MLEL)**2
      SUM0=SUM0+1./Z
      SUM1=SUM1+R(I)/Z
10    CONTINUE
      TMLES=DFLOAT(N)/2/SUM0/MLES**(1.5)
      MLES=TMLES**2
      MLEL=SUM1/SUM0
      ITER=ITER+1
      IF (ITER.GT.IMAX) GOTO 99
      IF (ABS(MLEL-MLELT).GT. .001*MLES)  GOTO 20
      IF (ABS(MLES-MLEST).GT. .05*MLES)   GOTO 20
99    RETURN
      END
C
C
      SUBROUTINE KSMIN (SSZE,RMLEL,RMLES,MKSL)
C
C
C     ***************************************************************
C     *                                                             *
C     * PURPOSE: TO PROVIDE THE LOCATION ESTIMATE THAT MINIMIZES*
C     *          THE KOLMOGOROV DISTANCE                            *
C     *                                                             *
C     * VARIABLES: DISA   - VERTICAL DISTANCE FROM THE CDF TO       *
C     *            DISB     THE EDF ABOVE/BELOW THE CDF             *
C     *            STEP   - THE INCREMENT ADDED ON TO OR SUB-       *
C     *                     TRACTED FROM THE PREVIOUS ESTIMATE      *
C     *            ITER   - THE NUMBER OF STEPS THAT ARE ADDED      *
C     *                     ON TO OR SUBTRACTED FROM THE INITIAL*
C     *                     ESTIMATE                                *
C     *            CKSDS  - KOLMOGOROV DISTANCE WITH THE MLE        *
```

```
C     *                        ESTIMATE AS THE PARAMETER VALUE      *
C     *              CKSL    - INITIAL PARAMETER VALUE              *
C     *              RKSDIS  - KOLMOGOROV DISTANCE AT EACH STEP     *
C     *              LKSDIS    TO THE RIGHT/LEFT OF THE INITIAL     *
C     *                        ESTIMATE                             *
C     *              EKSL    - PARAMETER VALUE AT EACH INCREMENT    *
C     *              RKSDS   - THE MINIMUM KOLMOGOROV DISTANCE TO   *
C     *              LKSDS     THE RIGHT/LEFT OF THE INITIAL        *
C     *                        ESTIMATE                             *
C     *              RKSL    - PARAMETER VALUE THAT YIELDS RKSDS/   *
C     *              LKSL      LKSDS                                *
C     *              J,K     - INTEGER VALUES TO DETERMINE THE      *
C     *                        ITERATION NUMBER THAT RKSDS/LKSDS    *
C     *                        OCCURS ON                            *
C     *              MKSDS   - MINIMUM KOLMOGOROV DISTANCE          *
C     *              MKSL    - PARAMETER VALUE THAT YIELDS MKSDS    *
C     *              RMLEL   - MLE ESTIMATE OF LOCATION PARAMETER   *
C     *              RMLES   - MLE ESTIMATE OF SCALE PARAMETER      *
C     *                                                             *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM      *
C     *                                                             *
C     *                                                             *
C     ***************************************************************
C
C
      REAL MKSDS,MKSL,LKSL,LKSDS,LKSDIS
      STEP=RMLES*SSZE
      ITER=1/SSZE
      CALL VEDIS (RMLEL,RMLES,DISA,DISB)
      IF (DISA.GT.DISB) THEN
         CKSDS=DISA
      ELSE
         CKSDS=DISB
      ENDIF
      CKSL=RMLEL
      RKSDS=CKSDS
      EKSL=CKSL
      RKSL=CKSL
5     J=0
      DO 10 I=1,ITER
        EKSL=EKSL+STEP
        CALL VEDIS (EKSL,RMLES,DISA,DISB)
        IF (DISA.GT.DISB) THEN
           RKSDIS=DISA
        ELSE
           RKSDIS=DISB
        ENDIF
        IF (RKSDIS.LT.RKSDS) THEN
          RKSDS=RKSDIS
          J=I
          RKSL=EKSL
        ENDIF
```

```
10    CONTINUE
      IF (J.EQ.ITER) GOTO 5
      EKSL=CKSL
      LKSDS=CKSDS
      LKSL=CKSL
15    K=0
      DO 20 I=1,ITER
        EKSL=EKSL-STEP
        CALL VEDIS (EKSL,RMLES,DISA,DISB)
        IF (DISA.GT.DISB) THEN
           LKSDIS=DISA
        ELSE
           LKSDIS=DISB
        ENDIF
        IF (LKSDIS.LT.LKSDS) THEN
          LKSDS=LKSDIS
          K=I
          LKSL=EKSL
        ENDIF
20    CONTINUE
      IF (K.EQ.ITER) GOTO 15
      IF (RKSDS.LT.LKSDS) THEN
         MKSDS=RKSDS
         MKSL=RKSL
      ELSE
         MKSDS=LKSDS
         MKSL=LKSL
      ENDIF
      RETURN
      END
C
C
      SUBROUTINE AEMIN (SSZE,RMLEL,RMLES,MAEL,IFLAG)
C
C
C     ***************************************************************
C     *                                                             *
C     * PURPOSE: TO PROVIDE THE LOCATION ESTIMATE THAT MINIMIZES*
C     *          THE CRAMER VON-MISES (CVM), ANDERSON-DARLING   *
C     *          (AD), OR THE WATSON DISTANCE                   *
C     *                                                         *
C     * VARIABLES: IFLAG  - FLAG USED TO DETERMINE WHICH DIS-   *
C     *                     TANCE MEASURE IS TO BE USED (CRAMER *
C     *                     VON MISES(0),ANDERSON-DARLING(1),   *
C     *                     WATSON(2))                          *
C     *            STEP   - THE INCREMENT ADDED ON TO OR SUB-   *
C     *                     TRACTED FROM THE PREVIOUS ESTIMATE  *
C     *            ITER   - THE NUMBER OF STEPS THAT ARE ADDED  *
C     *                     ON TO OR SUBTRACTED FROM THE INITIAL*
C     *                     ESTIMATE                            *
C     *            CAEDS  - CVM, AD, OR WATSON DISTANCE WITH THE*
C     *                     MLE ESTIMATE AS THE PARAMETER VALUE *
```

```
C     *               CAEL    - INITIAL PARAMETER VALUE                *
C     *               RAEDIS  - CVM, AD, OR WATSON DISTANCE AT EACH    *
C     *               LAEDIS    STEP TO THE RIGHT/LEFT OF THE          *
C     *                         INITIAL ESTIMATE                       *
C     *               EAEL    - PARAMETER VALUE AT EACH INCREMENT      *
C     *               RAEDS   - THE MINIMUM CVM, AD, OR WATSON         *
C     *               LAEDS     DISTANCE TO THE RIGHT/LEFT OF THE      *
C     *                         INITIAL ESTIMATE                       *
C     *               RAEL    - PARAMETER VALUE THAT YIELDS RAEDS/     *
C     *               LAEL      LAEDS                                  *
C     *               J,K     - INTEGER VALUES TO DETERMINE THE        *
C     *                         ITERATION NUMBER THAT RAEDS/LAEDS      *
C     *                         OCCURS ON                              *
C     *               MAEDS   - MINIMUM CVM, AD, OR WATSON DISTANCE    *
C     *               MAEL    - PARAMETER VALUE THAT YIELDS MAEDS      *
C     *               RMLEL   - MLE ESTIMATE OF LOCATION PARAMETER     *
C     *               RMLES   - MLE ESTIMATE OF SCALE PARAMETER        *
C     *                                                                *
C     *   NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM        *
C     *                                                                *
C     *                                                                *
C     ******************************************************************
      REAL MAEDS,MAEL,LAEDS,LAEDIS,LAEL
      STEP=RMLES*SSZE
      ITER=1/SSZE
      CALL AEDIS (RMLEL,RMLES,CAEDS,IFLAG)
      CAEL=RMLEL
      RAEDS=CAEDS
      RAEL=CAEL
      EAEL=CAEL
5     J=0
      DO 10 I=1,ITER
        EAEL=EAEL+STEP
        CALL AEDIS (EAEL,RMLES,RAEDIS,IFLAG)
        IF (RAEDIS.LT.RAEDS) THEN
          RAEDS=RAEDIS
          J=I
          RAEL=EAEL
        ENDIF
10    CONTINUE
      IF (J.EQ.ITER) GOTO 5
      EAEL=CAEL
      LAEL=CAEL
      LAEDS=CAEDS
15    K=0
      DO 20 I=1,ITER
        EAEL=EAEL-STEP
        CALL AEDIS (EAEL,RMLES,LAEDIS,IFLAG)
        IF (LAEDIS.LT.LAEDS) THEN
          LAEDS=LAEDIS
          K=I
          LAEL=EAEL
```

```
          ENDIF
 20     CONTINUE
        IF (K.EQ.ITER) GOTO 15
        IF (RAEDS.LT.LAEDS) THEN
          MAEDS=RAEDS
          MAEL=RAEL
        ELSE
           MAEDS=LAEDS
           MAEL=LAEL
        ENDIF
        RETURN
        END
C
C
        SUBROUTINE KUMIN (SSZE,RMLEL,RMLES,MKUL)
C
C
C       ***********************************************************
C       *                                                         *
C       * PURPOSE: TO PROVIDE THE LOCATION ESTIMATE THAT MINIMIZES*
C       *          THE KUIPER DISTANCE                            *
C       *                                                         *
C       * VARIABLES: DISA    - VERTICAL DISTANCE FROM THE CDF TO   *
C       *            DISB      THE EDF ABOVE/BELOW THE CDF        *
C       *            STEP    - THE INCREMENT ADDED ON TO THE      *
C       *                      PREVIOUS LOCATION ESTIMATE         *
C       *            ITER    - THE NUMBER OF STEPS THAT ARE ADDED *
C       *                      ON TO OR SUBTRACTED FROM THE INITIAL*
C       *                      ESTIMATE                           *
C       *            CKUDS   - KUIPER DISTANCE WITH THE MLE       *
C       *                      ESTIMATE AS THE PARAMETER VALUE    *
C       *            CKUS    - INITIAL PARAMETER VALUE            *
C       *            RKUDIS  - KUIPER  DISTANCE AT EACH STEP      *
C       *            LKUDIS    TO THE RIGHT/LEFT OF THE INITIAL   *
C       *                      ESTIMATE                           *
C       *            EKUS    - PARAMETER VALUE AT EACH INCREMENT  *
C       *            RKUDS   - THE MINIMUM KUIPER DISTANCE TO THE *
C       *            LKUDS     RIGHT/LEFT THE INITIAL ESTIMATE    *
C       *            RKUS    - PARAMETER VALUE THAT YIELDS RKUDS/ *
C       *            LKUS      LKUDS                              *
C       *            J,K     - INTEGER VALUES TO DETERMINE THE    *
C       *                      ITERATION NUMBER THAT RKUDS/LKUDS  *
C       *                      OCCURS ON                          *
C       *            MKUDS   - MINIMUM KUIPER DISTANCE            *
C       *            MKUS    - PARAMETER VALUE THAT YIELDS MKUDS  *
C       *            RMLEL   - MLE ESTIMATE OF LOCATION PARAMETER *
C       *                      (OR OTHER INITIAL ESTIMATE)        *
C       *            RMLES   - MLE ESTIMATE OF SCALE PARAMETER    *
C       *                      (OR OTHER INITIAL ESTIMATE)        *
C       *                                                         *
C       *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM  *
C       *                                                         *
```

```
      *                                                              *
C     ****************************************************************
C
      REAL LKUDIS,LKUDS,LKUL,MKUDS,MKUL
      STEP=RMLES*SSZE
      ITER=1/SSZE
      CALL VEDIS (RMLEL,RMLES,DISA,DISB)
      CKUDS=DISA+DISB
      CKUL=RMLEL
      RKUDS=CKUDS
      EKUL=CKUL
5     J=0
      DO 10 I=1,ITER
        EKUL=EKUL+STEP
        CALL VEDIS (EKUL,RMLES,DISA,DISB)
        RKUDIS=DISA+DISB
        IF (RKUDIS.LT.RKUDS) THEN
          RKUDS=RKUDIS
          J=I
          RKUL=EKUL
        ENDIF
10    CONTINUE
      IF (J.EQ.ITER) GOTO 5
      LKUDS=CKUDS
      LKUL=CKUL
      EKUL=CKUL
15    K=0
      DO 20 I=1,ITER
        EKUL=EKUL-STEP
        CALL VEDIS (EKUL,RMLES,DISA,DISB)
        LKUDIS=DISA+DISB
        IF (LKUDIS.LT.LKUDS) THEN
          LKUDS=LKUDIS
          K=I
          LKUL=EKUL
        ENDIF
20    CONTINUE
      IF (K.EQ.ITER) GOTO 15
      IF (RKUDS.LT.LKUDS) THEN
         MKUDS=RKUDS
         MKUL=RKUL
      ELSE
         MKUDS=LKUDS
         MKUL=LKUL
       ENDIF
89     RETURN
      END
      SUBROUTINE VEDIS (ELOC,ESCALE,DISA,DISB)
C
C
C     ****************************************************************
C     *                                                              *
```

```
C     *  PURPOSE: TO CALCULATE THE VERTICAL DISTANCE ABOVE AND  *
C     *           BELOW THE CDF TO THE EDF                      *
C     *                                                         *
C     * VARIABLES: ELOC   - LOCATION PARAMETER USED IN CDF      *
C     *            ESCALE - SCALE PARAMETER USED IN CDF         *
C     *            CAU(I) - ARRAY OF SAMPLE SIZE N USED TO STORE*
C     *                     CAUCHY CDF VALUES                   *
C     *            D1     - VERTICAL DISTANCE FROM THE CDF TO   *
C     *                     THE EDF WHERE THE CDF VALUE IS      *
C     *                     GREATER                             *
C     *            D2     - VERTICAL DISTANCE FROM THE EDF TO   *
C     *                     THE CDF WHERE THE EDF VALUE IS      *
C     *                     GREATER                             *
C     *                                                         *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM  *
C     *        OR HIGHER SUBROUTINE                             *
C     *                                                         *
C     *                                                         *
C     ***********************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL CAU(20)
      DO 5 I=1,N
         CAU(I)=.5+(1/PI)*(ATAN((R(I)-ELOC)/ESCALE))
5     CONTINUE
      DISB=CAU(1)
      DISA=ABS(CAU(1)-EDF(1))
      DO 10 I=2,N
         D1=ABS(CAU(I)-EDF(I-1))
         IF (D1.GT.DISB) THEN
            DISB=D1
         ENDIF
         D2=ABS(CAU(I)-EDF(I))
         IF (D2.GT.DISA) THEN
            DISA=D2
         ENDIF
10    CONTINUE
      RETURN
      END
C
C
      SUBROUTINE AEDIS (AEL,AES,DIS,IFLAG)
C
C
C     ***********************************************************
C     *                                                         *
C     *  PURPOSE: TO CALCULATE THE AREA BETWEEN THE CDF AND THE *
C     *           EDF                                           *
C     *                                                         *
C     * VARIABLES: AEL    - LOCATION PARAMETER USED IN CDF      *
C     *            AES    - SCALE PARAMETER USED IN CDF         *
```

```
C     *             CAU(I)  - ARRAY OF SAMPLE SIZE N USED TO STORE*
C     *                       CAUCHY CDF VALUES                    *
C     *             DIS     - AREA BETWEEN THE CDF AND EDF         *
C     *             IFLAG   - FLAG USED TO DETERMINE WHICH DIS-    *
C     *                       TANCE MEASURE IS TO BE USED (CRAMER *
C     *                       VON MISES)(0),ANDERSON-DARLING(1),   *
C     *                       WATSON(2)                            *
C     *             Z       - THE AVERAGE VALUE OF THE CAUCHY      *
C     *                       CUMULATIVE PROBABILITIES (USED IN    *
C     *                       WATSON'S DISTANCE MEASURE)           *
C     *                                                            *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM     *
C     *         OR HIGHER SUBROUTINE                               *
C     *                                                            *
C     *                                                            *
C     **************************************************************
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL CAU(20)
      DO 5 I=1,N
         CAU(I)=.5+(1/PI)*(ATAN((R(I)-AEL)/AES))
5     CONTINUE
      DIS=0
      IF (IFLAG.EQ.0) THEN
        DO 10, I=1,N
          DIS=DIS+(CAU(I)-(2*I-1)/(2*RN))**2
10      CONTINUE
        DIS=DIS+1/(12*RN)
      ENDIF
      IF (IFLAG.EQ.1) THEN
        DO 20, I=1,N
          DIS=DIS-(2*I-1)*(LOG(CAU(I))+LOG(1-CAU(N+1-I)))
20      CONTINUE
        DIS=DIS/RN-RN
      ENDIF
      IF (IFLAG.EQ.2) THEN
         Z=0
         DO 30, I=1,N
            DIS=DIS+(CAU(I)-(2*I-1)/(2*RN))**2
30       CONTINUE
         DIS=DIS+1/(12*RN)
         DO 40, I=1,N
            Z=Z+CAU(I)/RN
40       CONTINUE
         DIS=DIS-RN*(Z-.5)**2
      ENDIF
      RETURN
      END
```

Appendix C

Computer Program Listing of CSPEE

```
C     ************************************************************
C     *                                                          *
C     *                      PROGRAM                             *
C     *                                                          *
C     *    CAUCHY SCALE PARAMETER ESTIMATION AND                 *
C     *    EFFICIENCIES (CSPEE)                                  *
C     *                                                          *
C     *    WRITTEN BY CAPT JOHN O SOURS, AFIT/GSO-85D, FOR       *
C     *    MS THESIS                                             *
C     *                                                          *
C     *                   DECEMBER 1985                          *
C     *                                                          *
C     ************************************************************
C
C
      PROGRAM  CSPEE
C
C     ************************************************************
C     *                                                          *
C     *                                                          *
C     *  PURPOSE: TO COMPARE PARAMETER ESTIMATION TECHNIQUES OF *
C     *           THE SCALE PARAMETER OF THE TWO PARAMETER       *
C     *           CAUCHY DISTRIBUTION.  THE TECHNIQUES COMPARED *
C     *           ARE THE MAXIMUM LIKELIHOOD METHOD AND MINIMUM *
C     *           DISTANCE METHODS.                              *
C     *                                                          *
C     * VARIABLES: R(20)    - ARRAY WHICH HOLDS THE CAUCHY       *
C     *                       DEVIATES                           *
C     *            EDF(20)- ARRAY WHICH HOLDS VALUES OF THE      *
C     *                       EMPIRICAL DISTRIBUTION FUNCTION    *
C     *            PI       - 3.1415926 RADIANS                  *
C     *            N        - SAMPLE SIZE                        *
C     *            RN       - REAL VALUE FOR SAMPLE SIZE         *
C     *            DSEED    - SEED VALUE USED FOR GENERATING     *
C     *                       CAUCHY DEVIATES IN GGCAY           *
C     *            WK(60)  - ARRAY USED IN GGCAY FOR WORKSPACE   *
C     *                       (3 TIMES R(I))                     *
C     *            LOC      - LOCATION PARAMETER VALUE OF THE    *
C     *                       CAUCHY DISTRIBUTION FROM WHICH THE *
C     *                       DEVIATES ARE TAKEN                 *
C     *            SCALE    - SCALE PARAMETER VALUE OF THE       *
C     *                       CAUCHY DISTRIBUTION FROM WHICH THE *
C     *                       DEVIATES ARE TAKEN                 *
C     *            SSZE     - STEP SIZE USED IN THE MINIMUM      *
C     *                       DISTANCE TECHNIQUES                *
C     *            SUMME    - SUM OF THE MEAN SQUARED ERRORS     *
C     *            SUMMS      FOR THE MODIFIED SEMI-INTERQUARTILE *
C     *            SUMKU      RANGE, MLE ESTIMATE OF SCALE, KUIPER*
```

```
C     *               SUMWA    AND WATSON DISTANCE                          *
C     *               ESE    - SUMSE/SUMMS                                  *
C     *               EKU    - SUMKU/SUMMS                                  *
C     *               EWA    - SUMWA/SUMMS                                  *
C     *               DSE    - DIFFERENCE BETWEEN THE SCALE AND             *
C     *                        MODIFIED SEMI-INTERQUARTILE                  *
C     *                        RANGE                                        *
C     *               DSEQ   - DSE SQUARED (MEAN SQUARED ERROR)             *
C     *               DMS    - DIFFERENCE BETWEEN THE SCALE AND             *
C     *                        MLE OF SCALE                                 *
C     *               DMSQ   - DMS SQUARED (MEAN SQUARED ERROR)             *
C     *                        LOCATION                                     *
C     *               DKU    - DIFFERENCE BETWEEN THE SCALE AND             *
C     *                        KUIPER ESTIMATE OF SCALE                     *
C     *               DKUQ   - DKU SQUARED (MEAN SQUARED ERROR)             *
C     *               DWA    - DIFFERENCE BETWEEN THE SCALE AND             *
C     *                        WATSON ESTIMATE OF SCALE                     *
C     *               DWAQ   - DWA SQUARED (MEAN SQUARED ERROR)             *
C     *               MEDIAN - THE MEDIAN OF THE SAMPLE                     *
C     *               SEMIQ  - MODIFIED SEMI-INTERQUARTILE RANGE            *
C     *               MLEL   - MLE ESTIMATE OF LOCATION                     *
C     *               MLES   - MLE ESTIMATE OF SCALE                        *
C     *               KUL    - KUIPER ESTIMATE OF SCALE                     *
C     *               WAL    - WATSON ESTIMATE OF SCALE                     *
C     *                                                                     *
C     *                                                                     *
C     ***********************************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL WK(60),KUDIS,KUS,LOC,MLEL,MLES,MEDIAN
      DOUBLE PRECISION  DSEED
      PRINT*,'ENTER SAMPLE SIZE'
      READ*,N
      PRINT*,'ENTER # REPLICATIONS'
      READ*,NR
      PRINT*,'INPUT DSEED (0 FOR DEFAULT VALUE OF 4385673)'
      READ*,DSEED
      IF (DSEED.EQ.0) THEN
        DSEED=4385673.
      ENDIF
      PRINT*,'INPUT LOCATION AND SCALE PARAMETER'
      READ*,LOC,SCALE
      OPEN(UNIT=10,FILE='DATA',STATUS='NEW')
      WRITE(10,*)' '
      WRITE(10,*)'SAMPLE SIZE IS ',N,' LOC IS ',LOC,' SCALE IS
     *',SCALE
      WRITE(10,*)'  '
      PI=3.1415926
      SSZE=.01
      RN=N
      SUMSE=0
```

```
      SUMMS=0
      SUMKU=0
      SUMWA=0
C
C     CALCULATE EMPIRICAL DISRIBUTION FUNCTION
C
      DO 6 I=1,N
         EDF(I)=I/RN
6     CONTINUE
      DO 75 L=1,NR
C
C     GENERATE CAUCHY DEVIATES
C
      CALL GGCAY (DSEED,N,WK,R)
      DO 5 I=1,N
           R(I)=R(I)*SCALE+LOC
5     CONTINUE
C
C     ORDER MATRIX OF CAUCHY DEVIATES
C
      DO 25 I=1,N
         DO 35 J=1,N-1
         IF (R(J).GT.R(J+1)) THEN
            TEMP=R(J)
            R(J)=R(J+1)
            R(J+1)=TEMP
         ENDIF
35       CONTINUE
25       CONTINUE
C
C     COMPUTE MEDIAN AND MODIFIED SEMI-INTERQUARTILE RANGE
C
      CALL MEDSEM (MEDIAN,SEMIQ)
      DSE=SCALE-SEMIQ
      DSEQ=DSE**2
      SUMSE=SUMSE+DSEQ
C
C     CALCULATE MLE OF LOCATION AND SCALE
C
      CALL CMLE (MEDIAN,SEMIQ,MLEL,MLES)
C
      DMS=SCALE-MLES
      DMSQ=DMS**2
      SUMMS=SUMMS+DMSQ
C
C     MINIMIZE KUIPER DISTANCE TO ESTIMATE SCALE PARAMETER
C
      CALL KUMIN (SSZE,MLEL,MLES,KUS)
      DKU=SCALE-KUS
      DKUQ=DKU**2
      SUMKU=SUMKU+DKUQ
C
```

```
C     MINIMIZE THE WATSON DIS ANCE TO ESTIMATE THE SCALE PARAMETER
C
      CALL WAMIN (SSZE,MLEL,MLES,WAS)
      DWA=SCALE-WAS
      DWAQ=DWA**2
      SUMWA=SUMWA+DWAQ
75    CONTINUE
      ESE=SUMSE/SUMMS
      EKU=SUMKU/SUMMS
      EWA=SUMWA/SUMMS
      WRITE(10,*)'SUMSE = ',SUMSE
      WRITE(10,*)'SUMMS = ',SUMMS
      WRITE(10,*)'SUMKU = ',SUMKU
      WRITE(10,*)'SUMWA = ',SUMWA
      WRITE(10,*)' '
      WRITE(10,*)'ESE = ',ESE
      WRITE(10,*)'EKU = ',EKU
      WRITE(10,*)'EWA = ',EWA
      STOP
      END
C
C
      SUBROUTINE MEDSEM (MEDIAN,SEMIQ)
C
C     ****************************************************************
C     *                                                              *
C     *  PURPOSE: TO COMPUTE THE MEDIAN AND MODIFIED SEMI-INTER-*
C     *           QUARTILE RANGE OF THE CAUCHY RANDOM SAMPLE    *
C     *                                                              *
C     *                                                              *
C     *  VARIABLES: IH,IL,MH      -  INTEGER VARIABLES USED       *
C     *                              IN DETERMINING THE MODI-     *
C     *                              FIED SEMI-INTERQUARTILE      *
C     *                              RANGE                        *
C     *             D1,D2         -  DIFFERENCES BETWEEN THE      *
C     *                              MEDIAN AND CAUCHY RANDOM     *
C     *                              VARIATES                     *
C     *             SEMIQ1        -  VALUE USED IN AVERAGING      *
C     *                              THE RANGE IF THE SAMPLE      *
C     *                              SIZE IS EVEN AND EQUAL TO    *
C     *                              THE RANGE IF THE SAMPLE      *
C     *                              SIZE IS ODD                  *
C     *                                                              *
C     *  NOTE:  UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM *
C     *                                                              *
C     *                                                              *
C     ****************************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL MEDIAN
      MH=(N+2)/2
```

```
      MEDIAN=.5*(R(N-MH+1)+R(MH))
      SEMIQ=0
      IH=N
      IL=1
      DO 1 I=1,MH
      D1 = MEDIAN-R(IL)
      D2 = R(IH)-MEDIAN
      SEMIQ1=SEMIQ
      IF (D1.GT.D2) GOTO 2
      IH=IH-1
      SEMIQ=D2
      GOTO 1
2     IL=IL+1
      SEMIQ=D1
1     CONTINUE
      IF (MOD(N,2).EQ.0) SEMIQ=.5*(SEMIQ+SEMIQ1)
      RETURN
      END
C
C
      SUBROUTINE CMLE (MEDIAN,SEMIQ,MLEL,MLES)
C
C
C     ***********************************************************
C     *                                                         *
C     *  PURPOSE: TO CALCULATE THE MAXIMUM LIKELIHOOD ESTIMATORS*
C     *           OF THE LOCATION AND SCALE PARAMETERS FROM THE *
C     *           CAUCHY SAMPLE                                 *
C     *                                                         *
C     *                                                         *
C     * VARIABLES: IMAX   - MAXIMUM NUMBER OF ITERATIONS ALLOWED*
C     *            ITER   - ITERATION COUNTER                   *
C     *            MLELT  - VALUE OF MLE OF LOCATION/SCALE      *
C     *            MLEST    PARAMETERS AT THE BEGINNING OF EACH *
C     *                     ITERATION                           *
C     *            SUMO   - SUMMATION VARIABLES USED IN THE MLE *
C     *            SUM1     ITERATION PROCESS                   *
C     *            MLESSQ - MLES SQUARED                        *
C     *            Z      - VARIABLE USED IN THE MLE ITERATION  *
C     *                     PROCESS                             *
C     *            TMLES  - SQUARE ROOT OF THE MLE ESTIMATE OF  *
C     *                     SCALE                               *
C     *                                                         *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM  *
C     *                                                         *
C     *                                                         *
C     ***********************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL MLEL,MLES,MEDIAN,MLELT,MLEST,MLESSQ
      MLEL=MEDIAN
```

```
      MLES=SEMIQ
      IMAX=100
      ITER=0
20    MLELT=MLEL
      MLEST=MLES
      SUMO=O.
      SUM1=O.
      MLESSQ=MLES**2
      DO 10 I = 1,N
      Z=MLESSQ+(R(I)-MLEL)**2
      SUMO=SUMO+1./Z
      SUM1=SUM1+R(I)/Z
10    CONTINUE
      TMLES=DFLOAT(N)/2/SUMO/MLES**(1.5)
      MLES=TMLES**2
      MLEL=SUM1/SUMO
      ITER=ITER+1
      IF (ITER.GT.IMAX) GOTO 99
      IF (ABS(MLEL-MLELT).GT. .001*MLES)  GOTO 20
      IF (ABS(MLES-MLEST).GT. .05*MLES)   GOTO 20
99    RETURN
      END
C
C
      SUBROUTINE KUMIN (SSZE,RMLEL,RMLES,MKUS)
C
C
C     ************************************************************
C     *                                                          *
C     *  PURPOSE: TO PROVIDE THE SCALE ESTIMATE THAT MINIMIZES   *
C     *           THE KUIPER DISTANCE                            *
C     *                                                          *
C     * VARIABLES: DISA    - VERTICAL DISTANCE FROM THE CDF TO   *
C     *            DISB      THE EDF ABOVE/BELOW THE CDF         *
C     *            STEP1   - THE INCREMENT ADDED ON TO THE       *
C     *                      PREVIOUS SCALE ESTIMATE             *
C     *            STEP2   - THE INCREMENT SUBTRACTED FROM THE   *
C     *                      PREVIOUS SCALE ESTIMATE             *
C     *            ITER    - THE NUMBER OF STEPS THAT ARE ADDED  *
C     *                      ON TO OR SUBTRACTED FROM THE INITIAL*
C     *                      ESTIMATE                            *
C     *            CKUDS   - KUIPER DISTANCE WITH THE MLE        *
C     *                      ESTIMATE AS THE PARAMETER VALUE     *
C     *            CKUS    - INITIAL PARAMETER VALUE             *
C     *            AKUDIS  - KUIPER DISTANCE AT EACH STEP        *
C     *            BKUDIS    ABOVE/BELOW THE INITIAL ESTIMATE    *
C     *            EKUS    - PARAMETER VALUE AT EACH INCREMENT   *
C     *            AKUDS   - THE MINIMUM KUIPER DISTANCE         *
C     *            BKUDS     ABOVE/BELOW THE INITIAL ESTIMATE    *
C     *            AKUS    - PARAMETER VALUE THAT YIELDS AKUDS/  *
C     *            BKUS      BKUDS                               *
C     *            J,K     - INTEGER VALUES TO DETERMINE THE     *
```

```
C     *                      ITERATION NUMBER THAT AKUDS/BKUDS   *
C     *                      OCCURS ON                           *
C     *             MKUDS  - MINIMUM KUIPER DISTANCE             *
C     *             MKUS   - PARAMETER VALUE THAT YIELDS MKUDS   *
C     *             RMLEL  - MLE ESTIMATE OF LOCATION PARAMETER  *
C     *                      (OR OTHER INITIAL ESTIMATE)         *
C     *             RMLES  - MLE ESTIMATE OF SCALE PARAMETER     *
C     *                      (OR OTHER INITIAL ESTIMATE)         *
C     *                                                          *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM   *
C     *                                                          *
C     *                                                          *
C     ************************************************************
C
C
      REAL MKUDS,MKUS
      STEP1=RMLES*SSZE
      STEP2=.5*RMLES*SSZE
      ITER=1/SSZE
      CALL VEDIS (RMLEL,RMLES.DISA,DISB)
      CKUDS=DISA+DISB
      CKUS=RMLES
      AKUDS=CKUDS
      EKUS=CKUS
5     J=0
      DO 10 I=1,ITER
        EKUS=EKUS+STEP1
        CALL VEDIS (RMLEL.EKUS,DISA,DISB)
        AKUDIS=DISA+DISB
        IF (AKUDIS.LT.AKUDS) THEN
          AKUDS=AKUDIS
          J=I
          AKUS=EKUS
        ENDIF
10    CONTINUE
      IF (J.EQ.ITER) GOTO 5
      BKUDS=CKUDS
      BKUS=CKUS
      EKUS=CKUS
15    K=0
      DO 20 I=1,ITER
        EKUS=EKUS-STEP2
        IF (EKUS.LE.0) THEN
           GOTO 20
        ENDIF
        CALL VEDIS (RMLEL,EKUS,DISA,DISB)
        BKUDIS=DISA+DISB
        IF (BKUDIS.LT.BKUDS) THEN
          BKUDS=BKUDIS
          K=I
          BKUS=EKUS
        ENDIF
```

```
20    CONTINUE
      IF (K.EQ.ITER) GOTO 15
      IF (AKUDS.LT.BKUDS) THEN
         MKUDS=AKUDS
         MKUS=AKUS
      ELSE
         MKUDS=BKUDS
         MKUS=BKUS
      ENDIF
89    RETURN
      END
C
C
      SUBROUTINE WAMIN (SSZE,RMLEL,RMLES,MWAS)
C
C
C     ************************************************************
C     *                                                          *
C     *  PURPOSE: TO PROVIDE THE SCALE ESTIMATE THAT MINIMIZES   *
C     *           THE WATSON DISTANCE                            *
C     *                                                          *
C     * VARIABLES: STEP1  - THE INCREMENT ADDED ON TO THE        *
C     *                     PREVIOUS SCALE ESTIMATE              *
C     *            STEP2  - THE INCREMENT SUBTRACTED FROM THE    *
C     *                     PREVIOUS SCALE ESTIMATE              *
C     *            ITER   - THE NUMBER OF STEPS THAT ARE ADDED   *
C     *                     ON TO OR SUBTRACTED FROM THE INITIAL*
C     *                     ESTIMATE                             *
C     *            CWADS  - WATSON DISTANCE WITH THE MLE         *
C     *                     ESTIMATE,AS THE PARAMETER VALUE      *
C     *            CWAL   - INITIAL PARAMETER VALUE              *
C     *            AWADIS - WATSON DISTANCE AT EACH STEP         *
C     *            BWADIS   ABOVE/BELOW THE INITIAL ESTIMATE     *
C     *            EWAL   - PARAMETER VALUE AT EACH INCREMENT    *
C     *            AWADS  - THE MINIMUM WATSON DISTANCE          *
C     *            BWADS    ABOVE/BELOW THE INITIAL ESTIMATE     *
C     *            AWAL   - PARAMETER VALUE THAT YIELDS AWADS/   *
C     *            BWAL     BWADS                                *
C     *            J,K    - INTEGER VALUES TO DETERMINE THE      *
C     *                     ITERATION NUMBER THAT AWADS/BWADS    *
C     *                     OCCURS ON                            *
C     *            MWADS  - MINIMUM WATSON DISTANCE              *
C     *            MWAL   - PARAMETER VALUE THAT YIELDS MWADS    *
C     *            RMLEL  - MLE ESTIMATE OF LOCATION PARAMETER   *
C     *                     (OR OTHER INITIAL ESTIMATE)          *
C     *             RMLES  - MLE ESTIMATE OF SCALE PARAMETER     *
C     *                     (OR OTHER INITIAL ESTIMATE)          *
C     *                                                          *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM   *
C     *                                                          *
C     *                                                          *
C     ************************************************************
```

```
C
C
      REAL MWADS,MWAS
      STEP1=RMLES*SSZE
      STEP2=.5*RMLES*SSZE
      ITER=1/SSZE
      CALL AEDIS (RMLEL,RMLES,CWADS,IFLAG)
      CWAS=RMLES
      AWADS=CWADS
      AWAS=CWAS
      EWAS=CWAS
5     J=0
      DO 10 I=1,ITER
        EWAS=EWAS+STEP1
        CALL AEDIS (RMLEL,EWAS,AWADIS,IFLAG)
        IF (AWADIS.LT.AWADS) THEN
          AWADS=AWADIS
          J=I
          AWAS=EWAS
        ENDIF
10    CONTINUE
      IF (J.EQ.ITER) GOTO 5
      EWAS=CWAS
      BWAS=CWAS
      BWADS=CWADS
15    K=0
      DO 20 I=1,ITER
        EWAS=EWAS-STEP2
        IF (EWAS.LE.O) THEN
           GOTO 20
        ENDIF
        CALL AEDIS (RMLEL,EWAS,BWADIS,IFLAG)
        IF (BWADIS.LT.BWADS) THEN
          BWADS=BWADIS
          K=I
          BWAS=EWAS
        ENDIF
20    CONTINUE
      IF (K.EQ.ITER) GOTO 15
      IF (AWADS.LT.BWADS) THEN
         MWADS=AWADS
         MWAS=AWAS
      ELSE
         MWADS=BWADS
         MWAS=BWAS
      ENDIF
89    RETURN
      END
C
C
      SUBROUTINE VEDIS (ELOC,ESCALE,DISA,DISB)
C
```

```
C
C     ***********************************************************
C     *                                                         *
C     *  PURPOSE: TO CALCULATE THE VERTICAL DISTANCE ABOVE AND  *
C     *           BELOW THE CDF TO THE EDF                      *
C     *                                                         *
C     * VARIABLES: ELOC   - LOCATION PARAMETER USED IN CDF      *
C     *            ESCALE - SCALE PARAMETER USED IN CDF         *
C     *            CAU(I) - ARRAY OF SAMPLE SIZE N USED TO STORE*
C     *                     CAUCHY CDF VALUES                   *
C     *            D1     - VERTICAL DISTANCE FROM THE CDF TO   *
C     *                     THE EDF WHERE THE CDF VALUE IS      *
C     *                     GREATER                             *
C     *            D2     - VERTICAL DISTANCE FROM THE EDF TO   *
C     *                     THE CDF WHERE THE EDF VALUE IS      *
C     *                     GREATER                             *
C     *                                                         *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM  *
C     *         ORIGHER SUBROUTINE                              *
C     *                                                         *
C     *                                                         *
C     ***********************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL CAU(20)
      DO 5 I=1,N
         CAU(I)=.5+(1/PI)*(ATAN((R(I)-ELOC)/ESCALE))
5     CONTINUE
      DISB=CAU(1)
      DISA=ABS(CAU(1)-EDF(1))
      DO 10 I=2.N
         D1=ABS(CAU(I)-EDF(I-1))
         IF (D1.GT.DISB) THEN
            DISB=D1
         ENDIF
         D2=ABS(CAU(I)-EDF(I))
         IF (D2.GT.DISA) THEN
            DISA=D2
         ENDIF
10    CONTINUE
      RETURN
      END
C
C
      SUBROUTINE AEDIS (AEL,AES,DIS)
C
C
C     ***********************************************************
C     *                                                         *
C     *  PURPOSE: TO CALCULATE THE AREA BETWEEN THE CDF AND THE *
C     *           EDF                                           *
```

```
C     *                                                             *
C     * VARIABLES:  AEL     - LOCATION PARAMETER USED IN CDF        *
C     *             AES     - SCALE PARAMETER USED IN CDF           *
C     *             CAU(I)  - ARRAY OF SAMPLE SIZE N USED TO STORE*
C     *                       CAUCHY CDF VALUES                     *
C     *             DIS     - AREA BETWEEN THE CDF AND EDF          *
C     *             Z       - THE AVERAGE VALUE OF THE CAUCHY       *
C     *                       CUMULATIVE PROBABILITIES (USED IN     *
C     *                       WATSON'S DISTANCE MEASURE)            *
C     *                                                             *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM      *
C     *        OR HIGHER SUBROUTINE                                 *
C     *                                                             *
C     *                                                             *
C     ***************************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL CAU(20)
      DO 5 I=1,N
         CAU(I)=.5+(1/PI)*(ATAN((R(I)-AEL)/AES))
5     CONTINUE
      DIS=0
         Z=0
         DO 30, I=1,N
            DIS=DIS+(CAU(I)-(2*I-1)/(2*RN))**2
30       CONTINUE
         DIS=DIS+1/(12*RN)
         DO 40, I=1,N
            Z=Z+CAU(I)/RN
40       CONTINUE
         DIS=DIS-RN*(Z-.5)**2
      RETURN
      END
```

Appendix D

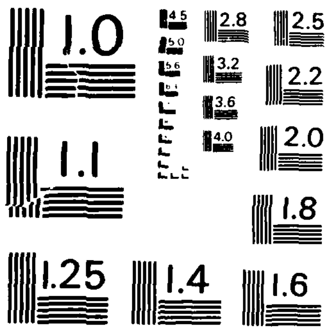
Computer Subprogram Listing of VEDIS and AEDIS for Censored Samples

```
C     ***********************************************************
C     *                                                         *
C     *                    SUBROUTINES                          *
C     *                                                         *
C     *        VEDIS AND AEDIS FOR CENSORED SAMPLES             *
C     *                                                         *
C     *        WRITTEN BY CAPT JOHN O SOURS, AFIT/GSO-85D, FOR  *
C     *        MS THESIS                                        *
C     *                                                         *
C     *                    DECEMBER 1985                        *
C     *                                                         *
C     ***********************************************************
C

      SUBROUTINE VEDIS (ELOC,ESCALE,DISA,DISB)
C
C
C     ***********************************************************
C     *                                                         *
C     *  PURPOSE: TO CALCULATE THE VERTICAL DISTANCE ABOVE AND  *
C     *           BELOW THE CDF TO THE EDF FOR CENSORED SAMPLES *
C     *                                                         *
C     * VARIABLES: ELOC   - LOCATION PARAMETER USED IN CDF      *
C     *            ESCALE - SCALE PARAMETER USED IN CDF         *
C     *            CAU(I) - ARRAY OF SAMPLE SIZE N USED TO STORE*
C     *                     CAUCHY CDF VALUES                   *
C     *            D1     - VERTICAL DISTANCE FROM THE CDF TO   *
C     *                     THE EDF WHERE THE CDF VALUE IS      *
C     *                     GREATER                             *
C     *            D2     - VERTICAL DISTANCE FROM THE EDF TO   *
C     *                     THE CDF WHERE THE EDF VALUE IS      *
C     *                     GREATER                             *
C     *                                                         *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM  *
C     *        OR HIGHER SUBROUTINE                             *
C     *                                                         *
C     *                                                         *
C     ***********************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL CAU(20)
      DO 5 I=3,N-2
         CAU(I)=.5+(1/PI)*(ATAN((R(I)-ELOC)/ESCALE))
5     CONTINUE
      DISB=ABS(CAU(3)-EDF(2))
      DISA=ABS(CAU(3)-EDF(3))
```

AD-A163 833   A COMPARISON OF ESTIMATION TECHNIQUES FOR THE TWO PARAMETER CAUCHY DISTRIBUTION(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGI . . J O SOURS   2/2

UNCLASSIFIED   DEC 85 AFIT/GSO/MA/85D-7   F/G 12/1   NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

```
      DO 10 I=4.N-2
         D1=ABS(CAU(I)-EDF(I-1))
         IF (D1.GT.DISB) THEN
            DISB=D1
         ENDIF
         D2=ABS(CAU(I)-EDF(I))
         IF (D2.GT.DISA) THEN
            DISA=D2
         ENDIF
10    CONTINUE
      RETURN
      END
C
C
      SUBROUTINE AEDIS (AEL,AES,DIS,IFLAG)
C
C
C     ************************************************************
C     *                                                          *
C     *  PURPOSE: TO CALCULATE THE AREA BETWEEN THE CDF AND THE *
C     *           EDF FOR CENSORED SAMPLES                       *
C     *                                                          *
C     * VARIABLES: AEL     - LOCATION PARAMETER USED IN CDF      *
C     *            AES     - SCALE PARAMETER USED IN CDF         *
C     *            CAU(I)  - ARRAY OF SAMPLE SIZE N USED TO STORE*
C     *                      CAUCHY CDF VALUES                   *
C     *            DIS     - AREA BETWEEN THE CDF AND EDF        *
C     *            IFLAG   - FLAG USED TO DETERMINE WHICH DIS-   *
C     *                      TANCE MEASURE IS TO BE USED (CRAMER *
C     *                      VON MISES)(0),ANDERSON-DARLING(1),  *
C     *                      WATSON(2)                           *
C     *            Z       - THE AVERAGE VALUE OF THE CAUCHY     *
C     *                      CUMULATIVE PROBABILITIES (USED IN   *
C     *                      WATSON'S DISTANCE MEASURE)          *
C     *                                                          *
C     *  NOTE: UNDEFINED VARIABLES ARE DEFINED IN MAIN PROGRAM   *
C     *        OR HIGHER SUBROUTINE                              *
C     *                                                          *
C     *                                                          *
C     ************************************************************
C
C
      COMMON R(20),EDF(20),PI,N,RN
      REAL CAU(20)
      DO 5 I=3,N-2
         CAU(I)=.5+(1/PI)*(ATAN((R(I)-AEL)/AES))
5     CONTINUE
      DIS=0
      IF (IFLAG.EQ.0) THEN
        DO 10, I=3,N-3
      DIS=DIS+(I**2/RN**2)*(CAU(I+1)-CAU(I))-(I/(RN))*(CAU(I+1)
     )     **2-CAU(I)**2)+(1/3.)*(CAU(I+1)**3-CAU(I)**3)
```

```
10      CONTINUE
       ENDIF
       IF (IFLAG.EQ.1) THEN
         DO 20, I=3,N-3
       DIS=DIS+(I**2/RN**2)*LOG(CAU(I+1)/(1-CAU(I+1)))+(I*2/RN)*
      )LOG(1-CAU(I+1))-CAU(I+1)-LOG(1-CAU(I+1))-(I**2/RN**2)*LOG
      )(CAU(I)/(1-CAU(I)))-(I*2/RN)*LOG(1-CAU(I))+CAU(I)
      )+LOG(1-CAU(I))
20      CONTINUE
       ENDIF
       IF (IFLAG.EQ.2) THEN
          Z=0
          DO 30, I=3,N-3
        DIS=DIS+(I**2/RN**2)*(CAU(I+1)-CAU(I))-(I/(RN))*(CAU(I+1)
      ) **2-CAU(I)**2)+(1/3.)*(CAU(I+1)**3-CAU(I)**3)
30        CONTINUE
          DO 40, I=3,N-3
          Z=Z+(I/RN)*(CAU(I+1)-CAU(I))-.5*(CAU(I+1)**2-CAU(I)**2)
40        CONTINUE
          Z=Z**2
          DIS=DIS-Z
       ENDIF
       RETURN
       END
```

VITA

John Orville Sours was born 23 April 1951 in Bellefontaine, Ohio. He graduated from Bellefontaine High School in 1969 and from Ohio State University in 1974. After enlisting in the United States Air Force in 1975, he was then selected for Officer's Training School in 1978 and commissioned a 2Lt in 1979. His first assignment was at the Cheyenne Mountain Complex in Colorado Springs where he served as a Surveillance Officer, Orbital Analyst, and Orbital Analyst Leader. After a remote tour at Clear AFS as the Chief of Space Object Identification, he was assigned to the Foreign Technology Division as a trajectories analyst prior to being selected to attend the Air Force Institute of Technology.

Permanent Address: 416 Louden St.
Urbana, Ohio 43078

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A163 833

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	UNCLASSIFIED
1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
3. DISTRIBUTION/AVAILABILITY OF REPORT	Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	AFIT/GSO/MA/85D-7
5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION	School of Engineering
6b. OFFICE SYMBOL (If applicable)	AFIT/MA
7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State and ZIP Code)	Air Force Institute of Technology Wright-Patterson AFB, OH 45433
7b. ADDRESS (City, State and ZIP Code)	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION	
8b. OFFICE SYMBOL (If applicable)	
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State and ZIP Code)	
10. SOURCE OF FUNDING NOS.	PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification)	See Box 19
12. PERSONAL AUTHOR(S)	John O. Sours, Captain, USAF
13a. TYPE OF REPORT	MS Thesis
13b. TIME COVERED	FROM ______ TO ______
14. DATE OF REPORT (Yr., Mo., Day)	1985 December
15. PAGE COUNT	90
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES		
FIELD	GROUP	SUB. GR.
12	01	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Estimates, Probability Distribution Functions, Distribution, Statistical Analysis, Statistical Inference, Statistical Distributions, Monte Carlo Method, Order Statistics

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Title: A COMPARISON OF ESTIMATION TECHNIQUES FOR THE TWO PARAMETER CAUCHY DISTRIBUTION

Thesis Advisor: Dr. Albert H. Moore, Professor

Approved for public release: IAW AFR 190-1.
Lynn Wolaver 16 Jan 86
LYNN E. WOLAVER
Dean for Research and Professional Development
Air Force Institute of Technology (AFC)
Wright-Patterson AFB OH 45433

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐
21. ABSTRACT SECURITY CLASSIFICATION	UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL	Dr. Albert H. Moore, Professor
22b. TELEPHONE NUMBER (Include Area Code)	513-255-2915
22c. OFFICE SYMBOL	AFIT/MA

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

Block 19. Abstract

Minimum distance estimators are compared to the maximum likelihood estimator of the location and scale parameters of the two parameter Cauchy distribution. Sample sizes of 6,8,10,12, and 16 are randomly drawn from a Cauchy distribution and used to try and improve upon the maximum likelihood estimates. 1000 samples for each sample size are generated and the mean squared error and relative efficiencies are computed. Comparison of the two methods is done by comparing the relative effici cies. Both censored and non-censored are used to compute the minimum distance est ıtes.

END
FILMED
3-86
DTIC